Title
Investigation of terminal alkene formation by acyl-CoA dehydrogenases in the biosynthesis of complex natural products

Author
Blake-Hedges, Jacquelyn

Publication Date
2019

Peer reviewed|Thesis/dissertation

Investigation of terminal alkene formation by acyl-CoA dehydrogenases in the biosynthesis of complex natural products

By

Jacquelyn M Blake-Hedges

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D Keasling, Co-chair
Professor Michelle Chang, Co-chair
Professor Evan Miller
Professor John Dueber

Fall 2019

Investigation of terminal alkene formation by acyl-CoA dehydrogenases in the biosynthesis of complex natural products

Abstract

Investigation of terminal alkene formation by acyl-CoA dehydrogenases in the biosynthesis of complex natural products

by

Jacquelyn M Blake-Hedges

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Jay D. Keasling, Co-chair

Professor Michelle C. Chang, Co-chair

Polyketides are a class of natural products known for their chemical complexity and bioactive properties. They are biosynthesized by an assembly line-like system termed polyketide synthases (PKSs), in which each module consists of various enzymatic domains, each of which has a single catalytic function. The polyketide is extended by two carbons and selectively reduced by each module. These biosynthetic principles make PKSs attractive engineering targets, because a simple swap of one enzymatic domain could change the final polyketide structure and effect an improved or novel bioactivity. However, gaping holes in our understanding of polyketide biosynthesis still exist. In particular, the activities of polyketide-associated enzymes remain relatively under-characterized and are generally not the primary focus of engineering efforts. These enzymes, which include glycosyl transferases, cytochrome P450s, and many others, often are responsible for imparting bioactive functional groups into the polyketide backbone or after the assembly of the "naked" polyketide. Many of these non-canonical functional groups discovered within polyketides are therefore the most desirable engineering targets. In this dissertation, I describe my efforts to characterize a family of terminal alkene-forming enzymes which were originally identified in a polyketide biosynthetic cluster. I show that the enzymes are regioselective and that regioselectivity is controlled by a shift in the protein structure. Finally, I show that these

enzymes are widespread in not only polyketide biosynthetic gene clusters, but also in other natural product clusters in the genomes of diverse Actinomycetes.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would first like to thank my parents for their constant and unwavering financial and emotional support. They have given me so much in life, more than I have ever asked for, and I will be forever indebted to them. As a child, they provided me with the ability to be curious and encouraged me to try anything and everything that piqued my interest. They inspired my passion for learning about and understanding natural phenomena: Dad, by taking me the explore the woods in our backyard to find salamanders and toads and identify plants; Mom, for helping me do science experiments in the kitchen and organizing excessively complicated science fair projects for a middle schooler. I do not think I would have pursued a graduate degree or even studied the sciences without their influence and inspiration. Finally, their financial support throughout graduate school has made my experience much more enjoyable. They gave me the privilege of focusing exclusively on my research, for which I will be forever grateful.

Next, I would like to thank my ever-supportive mentor Jay Keasling. He has created an amazing research environment for us that allowed me to be creative and to explore my passions. I cannot thank him enough for allowing me to learn through trial and error. Jay gave me ownership over my degree, and I am a much more independent and confident scientist because of it.

In addition to Jay, I have had numerous other mentors who were instrumental in my development as a scientist throughout my PhD. Leonard Katz served as a scientific advisor who constantly provided "big picture" guidance. Chris Petzold taught me so much about analytical chemistry and the importance of thorough experiment planning. Tristan de Rond was instrumental in my training and one of the biggest reasons I joined Keasling lab and quickly became independent. I had guidance from numerous other Keasling lab and JBEI members, including Jesus Barajas, Bob Haushalter, Maggie Brown, Andrew Hagen, Sean Poust, and many others.

I am also indebted to the many teachers and mentors I've had throughout my scientific journey. In particular, my high school teachers Mrs. Schuck, Mrs. Reister, and Mr. Shaw inspired my academic scientific journey. In college, I had wonderful teacher-mentors in the William and Mary Chemistry department in Elizabeth Harbron and Jonathan Scheerer. My love of proteomics was inspired by a lab class I took with JC Poutsma and Kurt Williamson. I was also fortunate to participate in a summer internship program at the EPA's analytical chemistry lab where I learned skills I used throughout my PhD thanks to my mentor Alaa Kamel. Most importantly, I am forever grateful to Kristin

Wustholz for her mentorship and advice regarding both research and advice. She gave me freedom to try new things in lab which inspired my choice of graduate lab. Most importantly, she gave me the confidence that I could succeed in any scientific endeavor, even if that meant completely switching to a new subject area in graduate school.

There have been many others who have supported me throughout this process. In particular, I want to thank my scientific partner-in-crime Mitch Thompson for including me in his exciting enzymology projects and generally for being a great friend throughout grad school. Others who have been influential in my graduate school career include Alberto Rodriguez, Connie Bailey, Henrique Pereira, Pablo Cruz-Morales, Maren Wehrs, Sam Curran, Jesus Barajas, Bob Haushalter, Yan Chen, Leanne Chan, Edward Baidoo, Anna Lechner, Ryan Phelan, Chris Wheeler, Kat Saxton, Alisha Contractor, Sam Keyser, Rebecca Durr, and Katie Barry.

Finally, I would like to thank my undergraduate mentees who all helped with various aspects of my graduate work: Jeffrey Chen, Danika Nimlos, Rohith Krishna, Catalina Alonso-Martinez, and Ronald Kam. By working with them, I learned so much more about my own research and how to be a better mentor. I am so proud of their accomplishments so far and cannot wait to see where the future takes them.

Chapter 1. Introduction and dissertation organization

1.1. Polyketide biosynthesis as a source of undiscovered enzymology

Polyketides are a class of natural products known for their chemical complexity and bioactive properties. They are biosynthesized by an assembly line-like system termed polyketide synthases (PKSs), in which each module consists of various enzymatic domains, each of which has a single catalytic function. The polyketide is extended by two carbons and selectively reduced by each module. These biosynthetic principles make PKSs attractive engineering targets, because a simple swap of one enzymatic domain could change the final polyketide structure and effect an improved or novel bioactivity.

However, gaping holes in our understanding of polyketide biosynthesis still exist. In particular, the activities of polyketide-associated enzymes remain relatively under-characterized and are generally not the primary focus of engineering efforts. These enzymes, which include glycosyl transferases, cytochrome P450s, and many others, often are responsible for imparting bioactive functional groups into the polyketide backbone or after the assembly of the "naked" polyketide. Many of these non-canonical functional groups discovered within polyketides are therefore the most desirable engineering targets. In this dissertation, I describe my efforts to characterize a family of terminal alkene-forming enzymes which were originally identified in a polyketide biosynthetic cluster. I show that the enzymes are regioselective and that regioselectivity is controlled by a shift in the protein structure. Finally, I show that these enzymes are widespread in not only polyketide biosynthetic gene clusters, but also in other natural product clusters in the genomes of diverse Actinomycetes.

1.2. Dissertation organization

In Chapter 2 I introduce polyketide synthases and describe the fundamental principles of polyketide biosynthesis. I focus on polyketide synthase engineering strategies, both at the protein and organism level. In the protein engineering section, engineering strategies for each polyketide synthase domain, for loading modules, and for intermodular linkers are discussed. In the host engineering section, I highlight several engineering strategies including precursor-directed biosynthesis, precursor pool engineering, expression of phosphopantetheinyl transferases, and transcriptional regulation/gene refactoring. Finally, I finish by describing how protein and host engineering methods can be combined to synergistically improve polyketide biosynthesis.

In **Chapter 3** I describe the structural and biochemical characterization of a terminal alkene-forming acyl-CoA dehydrogenase, tcsD, that is implicated in the biosynthesis of the allylmalonyl-CoA extender unit from the biosynthesis of the immunosuppressant polyketide FK506. First, I disprove an earlier hypothesis that two enzymes from the allylmalonyl-CoA biosynthetic pathway can act interchangeably; instead, the pathway proceeds in a linear fashion. I then show that tcsD is regioselective for the γ,δ-position of acyl-ACP substrates and that it only acts on α,β-unsaturated substrates. By solving the crystal structure of the enzyme, I show that this regioselectivity is mediated by a lateral shift in the positioning of the FAD cofactor within the enzyme. Finally, I identify homologs of the enzyme and show that a phylogenetically distant homolog catalyzes the same reaction as tcsD and is also regioselective for the γ,δ-position of substrates.

In **Chapter 4** I have added supplementary information and discussion point related to chapter 3. This includes discussions regarding LC-MS/MS method development for monitoring enzymatic reactions on carrier protein-bound intermediates and the presence of an unknown metabolite that crystallized in the tcsD active site. In addition, I have included several figures that support the claims made in chapter 3.

Chapter 2. Engineered polyketides: synergy between host and protein engineering

Including material from published work: Barajas, J.F.,* **Blake-Hedges, J.M.**,* Bailey, C.B.,* Curran, S., Keasling, J.D. Engineered polyketides: synergy between host and protein engineering, *Synthetic and Systems Biotechnology*, **2017**, 2, 147-166.
*authors contributed equally

2.1. Abstract

Metabolic engineering efforts toward rewiring metabolism of cells to produce new compounds often require the utilization of non-native enzymatic machinery that is capable of producing a broad range of chemical functionalities. Polyketides encompass one of the largest classes of chemically diverse natural products. With thousands of known polyketides, modular polyketide synthases (PKSs) share a particularly attractive biosynthetic logic for generating chemical diversity. The engineering of modular PKSs could open access to the deliberate production of both existing and novel compounds. In this review, we discuss PKS engineering efforts applied at both the protein and cellular level for the generation of a diverse range of chemical structures, and we examine future applications of PKSs in the production of medicines, fuels and other industrially relevant chemicals.

2.2. Introduction

Polyketides are one of the largest classes of natural products, possessing immense structural diversity and complex chemical architectures. Many polyketides (PKs) are among the most important secondary metabolites for their applications in medicine, agriculture, and industry. Examples include anticancer drugs (epothilone),[1,2] antibiotics (erythromycin),[3] insecticides (spinosyn A)[4] and antifungals (amphotericin B).[5] These particular examples of polyketides are biosynthesized by multimodular enzyme complexes known as type I modular polyketide synthases (PKSs). Working in an assembly-line fashion, multimodular PKSs assemble and tailor readily available acyl-CoAs within the host cell into large, complex, chiral molecules.[6] Each of these PKSs comprises a series of modules that can be further dissected into a series of domains responsible for the extension of the polyketide backbone through condensation and selective reductive processing of an acyl-CoA building block. The collinear architecture of these modules, apparent by inspection of the domains present and the predictive selectivity motifs harbored within, provide insights into the chemical connectivity and stereochemical configuration of the polyketide metabolite from analysis of its coding sequence.

One of the best-studied PKSs is the 6-deoxyerythronolide synthase (DEBS) (**Figure 2-1A**), which is responsible for synthesizing the macrolactone core of the antibiotic, erythromycin.[7] The catalytic domains of DEBS are expressed within modules that are each responsible for a single round of chain elongation and reductive processing. To this end, the loading acyltransferase (AT) domain loads the acyl carrier protein (ACP) with a starter unit derived from propionyl-CoA (**Figure 2-1A**). The ketosynthase (KS) within each module catalyzes decarboxylative carbon-carbon bond formation between an acyl precursor and the ACP-bound methylmalonyl derivative. Unlike fatty acid synthases (FASs), the occurrence of reductive domains within modules varies, and PKS intermediates typically exhibit various levels of reduction. If present, the ketoreductase (KR) converts the β-ketone to an alcohol using NADPH. The dehydratase (DH) eliminates the alcohol to form an olefin, and the enoylreductase (ER) utilizes NADPH to reduce the olefin to a methylene. Finally, a thioesterase (TE) domain, located at the terminal of DEBS 3 module, catalyzes the release and cyclization to produce the macrolactone, 6-deoxyerythronolide (6-dEB). The structure and mechanism of each PKS domain is reviewed in detail elsewhere.[8–12]

With this collinear biosynthetic logic in mind, engineered PKSs have the potential to become an effective retrobiosynthetic platform to produce molecules that are difficult or too complex to acquire via traditional synthetic means (**Figure 2-1B-C**). From DNA sequence, one could control chemical structure by successfully modifying and rearranging existing polyketide modules and domains.[13,14] Moreover, rationally-designed PKSs could be introduced into a variety of engineered hosts[15–17] capable of expressing these large PKS complexes while providing the necessary precursor metabolites to biosynthesize a target chemical. In this review, we highlight PKS engineering efforts at both the protein level and the host/cellular level. We further aim to describe PKS engineering efforts within the context of metabolic engineering and introduce the idea of successful PKS/host modifications for both traditional medicinal applications as well as the production of fuels and commodity chemicals.

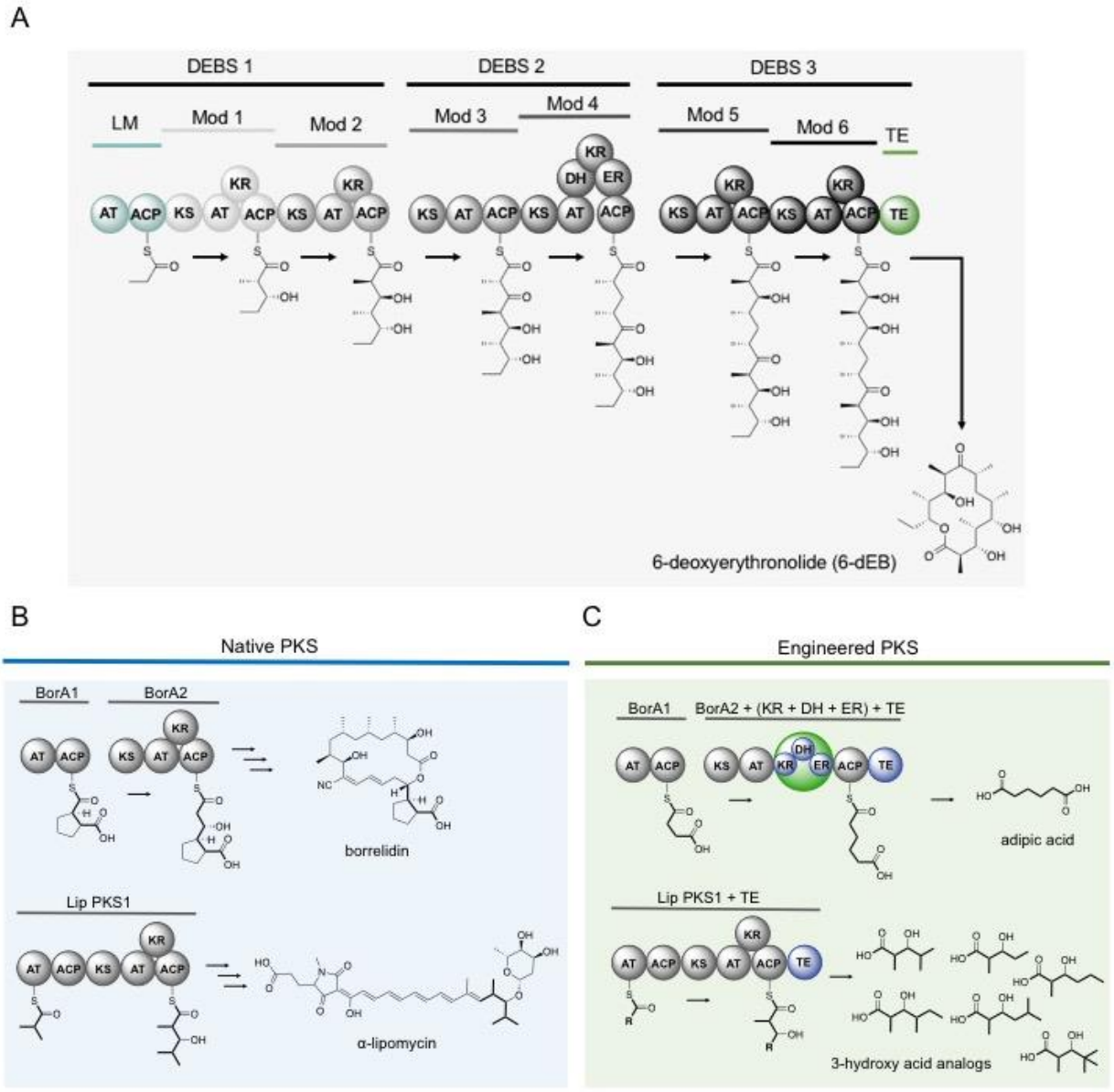


Figure 2-1 Biosynthesis of 6-deoxyerythronolide and examples of both native and engineered polyketide synthases. A) Modular biosynthesis of 6-deoxyerythronolide by the well-studied 6-deoxyerythronolide polyketide synthase. **B)** The carboxylic acid starter unit promiscuity by the borrelidin PKS was utilized to produce adipic acid (**C**) from succinic acid and malonic acid using an engineered BorA2 containing a full reductive loop (highlighted in green circle) and a thioesterase. **C)** The broad starter unit selectivity of the lipomycin PKS was utilized to produce 3-hydroxyacid congeners from branched acids and methylmalonyl-CoA (**C**). The carbon backbone for both the native and engineered polyketides are represented in bold. Grey domains represent the native pathway and blue domains represent engineered insertions.

2.3. PKS Protein Engineering

The ability to tailor the molecular architecture of polyketide metabolites through the inclusion of various reductive domains or domains with altered selectivity has long been the promise of PKSs as a retrobiosynthetic platform. In this section, we discuss the current knowledge of PKS engineering at the protein level. We have divided the PKS protein engineering section into sections based on domain type. Within the types of domains, we have selected the most engineerable targets. We will not focus on KS or ACP domain engineering in this review, as they are arguably the least targetable domains based on the chemistry and functions they perform, respectively. In addition, methyltransferase domains, which transfer an S-adenosyl-methionine-derived methyl group to the α-carbon of the β-keto intermediate, are somewhat rare and less well characterized, and thus will not be discussed here. In each section, we will first give a basic overview of the current state of knowledge regarding the specific domain(s) in question. Next, we will highlight some significant accomplishments in engineering, both via site-directed mutagenesis and/or domain swapping experiments. Because of the extensive amount of published PKS research, we cannot include all examples of PKS engineering within the scope of this review. Nevertheless, numerous representative examples are highlighted.

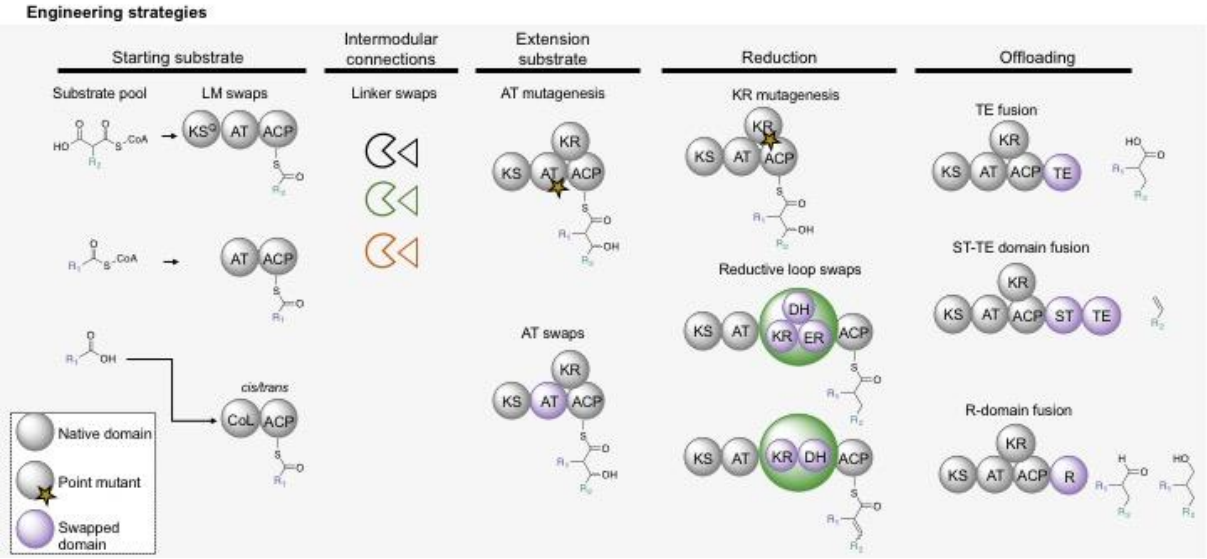


Figure 2-2 Summary of polyketide synthase engineering strategies highlighted in this chapter. A) Starter unit selectivity and incorporation is mediated by either native or non-native swapped LMs. Intermolecular linker regions allow for successful communications between domains. Chemical diversity is further increased by varying the extender building blocks. AT mutagenesis or AT swaps mediate incorporation of various extender units into a polyketide intermediate. Various degrees of reduction at the β-keto position can be accomplished by KR mutagenesis, KR swaps and/or the insertion of full or partially full reductive loops,

containing DH and ER domains. Release of the polyketide intermediate is mediated by various releasing domains, which further increase chemical diversity into the final product.

2.3.1. Loading modules

Nature has evolved several mechanisms for activating acyl substrates to initiate PK biosynthesis. To begin chain formation, modular type I PKSs employ a loading module (LM) to select the priming unit. LMs are categorized based on their domain architecture and the mechanism by which each activates substrates to begin chain formation. Although LMs are not officially characterized within the field, for simplicity within this review we will refer to each class of LMs with a representative letter (e.g. "type A" or "A-type") so as not to confuse the reader with the type I, II, or III PKS designations used to describe the entire assembly line systems.

The most common LM organization consists of a condensation-incompetent KS^Q (named for the active site C→ Q mutation), AT, and ACP domain (Type A LM). In type A LMs, the KS^Q decarboxylates malonyl- or methylmalonyl-CoA to yield acetyl- or propionyl starter units, respectively (**Figures 2-2, 2-3**).[18] The AT domains of these modules are strictly specific for CoA esters of dicarboxylic acids (malonyl- or methylmalonyl-CoA)[18] and share a conserved arginine residue with extender AT domains that is used to stabilize the free carboxyl moiety of the substrate.[19] Type B LMs prime polyketide biosynthesis with a much broader range of substrates (**Figures 2-2, 2-3**), and their domain organization consists of an AT and ACP didomain. The AT domain selects a CoA-bound priming unit and transfers it to its cognate ACP where it is poised for transfer to the KS of the first extension module. Because type B LM ATs are not limited to utilizing β-carboxy-CoA starters, they tend to recognize a more diverse set of starter units derived from other acyl-CoAs. For example, the B-type avermectin LM natively primes with either 2-methylbutyryl-CoA or isobutyryl-CoA, and it can also accept a large number of other substrates.[20] Similarly, while the related lipomycin LM primes biosynthesis with isobutyryl-CoA *in vivo*, *in vitro* it also loads a variety of other branched fatty acyl-CoAs (**Figure 2-1 B-C**).[21] Nevertheless, type B domain architecture does not always guarantee promiscuity. Another B-type LM from the borrelidin PKS is selective for dicarboxylic acid starter units both *in vivo*[22] and *in vitro* (**Figure 2-1 B-C**).[23]

The final common organization (type C) of LMs consists of a CoA-Ligase-type (CoL) domain located upstream of an ACP, which activates a carboxylic acid substrate in an ATP-dependent fashion so that it can then be loaded onto the ACP (**Fig. 2,3**). While some LMs have the CoL domain *in cis* (e.g. rifamycin,[24] rapamycin[25]), others have a separately encoded CoL domain that activates the ACP of the loading domain *in trans* in

a mechanistically similar fashion (e.g. aureothin,[26] and several mycobacterial polyketide synthases[27]). It is also common to find other accessory domains within type C LMs, such as the enoylreductase domains observed in FK506 and rapamycin biosynthesis.[25,28]

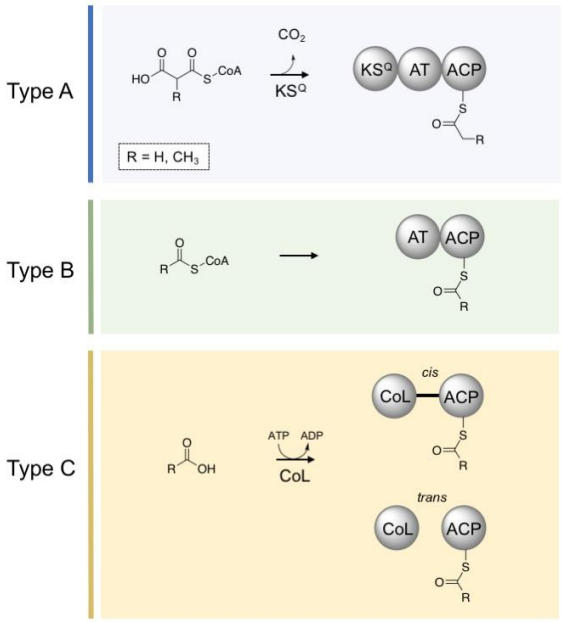


Figure 2-3 Examples of loading modules. In type A LMs, the KS^Q decarboxylates malonyl- or methylmalonyl-CoA to yield acetyl- or propionyl starter units, respectively. Type B LMs consist of an AT which selects a CoA-bound priming unit and transfers it to its cognate ACP. Type C LMs consists of a CoL domain located upstream of an ACP. The CoL activates a carboxylic acid substrate in an ATP-dependent fashion in order to load it onto the ACP, either in *cis* or in *trans*. It is also common to find other accessory domains within type C LMs, between the CoL and ACP domain (represented by the black line between the *cis* CoL and ACP).

The inherent promiscuity of certain loading modules and the wide variety of potential starter units make the initial precursor loading process an attractive protein engineering target. To date, most LM engineering efforts have focused on swapping LMs with type A or type B architectures that incorporate short fatty acyl groups into polyketides. Possibly the first LM swap successfully placed the tylosin LM (type A) into the platenolide (type A LM) biosynthetic pathway, resulting in the production of 16-methyl platenolide derived from propionyl-CoA (methylmalonyl-derived) instead of the native acetyl-CoA (malonyl-derived) starter unit.[29] In a similar experiment, Leadlay and coworkers exchanged the native 6-deoxyerythronolide LM (type B) for the avermectin LM (type B), increasing the diversity of erythromycins produced *in vivo*.[30] Both of these

early examples demonstrated the feasibility of swapping loading modules with consistent domain architectures.

Mixed LM type swaps have also been explored. By replacing the type B LM of DEBS M1+TE with the type A LMs from the oleandomycin or tylosin pathways, production of triketide lactones (TKLs) derived from almost exclusively acetate- or propionate-derived starter units, respectively, was achieved.[18] Under the same conditions, the native type B DEBS LM architecture lends itself to broader starter unit selectivity, resulting in the production of both types of TKLs.[31] Therefore, by changing LM type, greater starter unit fidelity was achieved. In addition, an AT swap of the oleandomycin (type A LM) loading AT with the extending AT from the second rapamycin module (rapAT2, also conferring malonyl-CoA specificity) resulted in production of C13-methyl erythromycins.[18] A similar AT swap into the nystatin LM *nysA* (type A) resulted in the restoration of nystatin production in a *nysA* knockout strain of *S. noursei*, albeit at much lower titers than the native system. In addition, swapping in a methylmalonyl-CoA-specific AT domain resulted in no production of the nystatin analog, perhaps due to the innate specificity of the downstream *nysB* module or the *nysA* KS^Q domain.[32] These swaps illustrate the potential to exercise tighter control over starter units incorporated into polyketides that normally begin with type B LMs in order to engineer systems to precisely control the metabolites produced.

LM swaps involving type C LMs could provide mechanisms for introducing even greater diversity into polyketides than type B LMs based on the range of starter units the type C LMs accept. Dihydroxycyclohexene-carboxoyl-CoA,[33] benzoyl-CoA,[34] and 3-amino-5-hydroxybenzyl-CoA[24] are just a few of the starter units incorporated via this type of LM. Attempts to switch the type C LM of the rimocidin PKS protein (RimA) with the type A LM from the nystatin PKS (NysA) showed that while RimA could initiate nystatin biosynthesis in the native *S. noursei* producer, *nysA* could not rescue rimocidin biosynthesis in a *rimA* knockout strain of *S. diastaticus*. In addition, RimA natively accepts both acetate and butyrate starter units, but only acetate units could be incorporated into the nystatin skeleton, suggesting that the downstream nys PKS is gatekeeping.[35] However, a swap of the non-canonical soraphen LM SorA (ACP-KS-AT-AT architecture) into DEBS1+TE produced a small amount of the expected benzoyl-CoA-derived TKL.[36] In addition, a swap of the first SorA AT domain into the DEBS LM also resulted in the production of a TKL incorporating benzoyl-CoA. These experiments illustrate that more exotic LM swaps, including type C LM swaps and domain swaps within LMs, may be tolerated by modular PKSs. However, incorporation of non-native starter units via noncognate LMs could require additional engineering of

the downstream modules of the polyketide synthase, as they might not accept or act on unusual or bulky functionalities.[37]

Currently, there is no flawless method for LM engineering to incorporate nonnative starter units into polyketides. The most common approach involves swapping a full LM (that selects for a desired starter) in place of the native LM. This has been met with limited success, and a more detailed understanding of where bottlenecks arise is needed before LM swaps can become a more common practice. Because LMs can occur *in cis* or *in trans*, the native protein-protein interactions that occur within each type of system should be considered. It might be best to retain the native intermodular linker for *in cis* LMs, and in the same way, LMs used to complement *in trans* may be more successful if domains are engineered with C-terminal linker domains from the native system (**Figure 2-2**). Other possibilities for retaining native protein-protein interactions include performing AT swaps within LMs instead of swapping full modules. We hypothesize that more conservative swaps (in terms of chemical structure) will be more successful because of the gatekeeping functions of the downstream PKS modules. Interactions between the LM ACP and the first KS of the downstream extension module should also be considered, as improper protein-protein interactions will prevent chain translocation to the first extension module. However, until a more detailed and systematic study of LM swaps is published, the importance of each of these factors for each unique system is unknown. Therefore, with the current state of knowledge, each LM engineering attempt should be optimized individually and should be attempted with multiple domain boundaries in order to find the most productive system.

2.3.2. Acyltransferase domains

The acyltransferase domain is responsible for selecting the CoA-based starter and/or extender units that form the majority of the carbon backbone of the growing polyketide. AT domains are attractive engineering targets for several reasons. First, we can easily predict the substrate specificity of most AT domains based on conserved residues. Secondly, AT domains are the primary sources of diversification at the α-carbon, a diversity that extends beyond the extent of the various oxidations states of the β-keto groups to include other heteroatoms, halogens, and unique functionalities that are otherwise unachievable via traditional polyketide chemistry.[38–43] Finally, AT domains are naturally more promiscuous in comparison to other PKS domains,[38,44,45] making precursor- directed mutagenesis viable. For these reasons, AT domains are the most well-studied of all PKS domains in terms of engineering via site-directed mutagenesis, domain swapping, and other host-level techniques (*vide infra*).

While nature uses numerous diverse extender units to form PKSs, the majority of known ATs select for either malonyl- or methylmalonyl-CoA.[46] Well before the first KS-AT didomains structures were solved for DEBS M3[47] and M5,[48] consensus sequences were identified within the AT domain that correspond to malonyl-(HAFH) or methylmalonyl-CoA (YASH) specificity.[49] With structural information in hand, identification of active site residues for targeted mutagenesis became much easier. Later on, more diverse extender units were identified, revealing other specificity motifs in the associated AT domains. The AT from module 3 of the epothilone cluster, for example, accepts both malonyl- and methylmalonyl-CoA and contains a "HASH" motif,[50] while the allylmalonyl-CoA-specific AT from module 4 of the FK506 cluster has a "CPTH" motif at the same location.[51] Several other ATs with unique specificity-conferring residues at this position have also been discovered and their cognate extender units identified, significantly expanding the chemistries accessible via PKS-mediated biosynthesis.[46,52] Nevertheless, simply mutating these residues to the appropriate specificity motif does not necessarily confer specificity to the desired extender unit. Mutation of the of the methylmalonyl-specific "YASH" motif to "HAFH" in DEBS AT1,[45] AT4, and AT6[53] increased promiscuity, allowing the domains to accept and transfer both malonyl- and methylmalonyl-CoA. However, later *in vitro* investigations demonstrated that these modules had greatly reduced activity.[54] Mutations outside the active site can also lead to more promiscuous AT domains, although a greater amount of retention of native substrate specificity is frequently observed.[53,55,56] These types of mutations can even abolish the transacylation activity of the domain altogether.[51] Evidence suggests that the specificity conferred by these conserved residues could be overcome if nonnative extender unit concentrations are high enough relative to the native extender.[57]

The incorporation of rare extender units can also be achieved by targeting other residues with site-directed mutagenesis (**Figure 2-2**). For example, mutating an active site valine to an alanine in DEBS AT6 created a more promiscuous enzyme that was able to incorporate various nonnative extender units such as propargylmalonyl-, allylmalonyl-, and ethylmalonyl-CoA into the erythromycin backbone.[56] More recent work has revealed that targeted active site mutagenesis screens can be more effective at producing ATs that are more specific for nonnative extenders than the native substrate. Williams and coworkers showed that DEBS AT6 could be converted to a propargylmalonyl-CoA-specific domain by screening a library of active site mutants and identifying a Y → R mutation that changes the native specificity.[43] Combined with exogenous feeding of precursors (*vide infra*), AT mutagenesis screens could be used to create diverse libraries of polyketides from just one engineered PKS.

AT swapping is also commonly undertaken as a method for constructing hybrid modules capable of producing new polyketides with novel α-carbon substituents (**Figure 2-2**). Traditionally, AT swaps were constructed via restriction enzyme-based cloning, thus relying on conserved restriction sites flanking the AT domain. One of the first attempts at a full AT swap was performed by Leadlay and coworkers in which the native AT domain from DEBS1-TE (malonyl-CoA specific) was replaced with the AT domain from module 2 of the rapamycin synthase (methylmalonyl-CoA specific).[58] The hybrid protein was able to produce TKLs exclusively derived from methylmalonyl-CoA with little to no effect on yield, highlighting the significant potential of the technique for producing novel polyketides.

Katz and coworkers showed that, when placed in the context of a full modular PKS system, AT domain exchanges are not as efficient as the native system.[39] Swaps of malonyl-CoA-specific ATs from the pikromycin and rapamycin PKSs into DEBS modules 1 and 2, in particular swaps into the second module, resulted in significantly lower titers of the expected erythromycin derivatives.[40–42] In a similar study, a swap of the second rapamycin AT into module 6 of DEBS was capable of producing 25 mg/L of the appropriate 6-deoxyerythronolide B analogs, suggesting that the efficiency of AT-swapped modules may be affected by the position of the module within the larger context of the assembly line.[59] However, a successful swap of the ethylmalonyl-CoA-specific AT domain from niddamycin in module 4 of DEBS, suggested that other factors may be at play.[60] A functional rapamycin AT2 swap into DEBS module 4 finally illustrated that the domain boundaries for AT swaps (and any domain swap) may need to be optimized for each unique acceptor module-AT pairing.[61] Several other examples of AT domain exchanges emerged, each reporting extremely low titers due to the engineered nature of the systems at hand.[62,63] DEBS module 4 was the last module in the model system to be successfully engineered in this manner by varying the domain boundaries used, suggesting that modules with full reductive loops may necessitate different AT swap domain boundaries because of the difference in the domains and linkers that surround the AT.

The mixed success observed with AT swaps led to the pursuit of a more fundamental understanding of the mechanisms of catalysis in native versus engineered AT systems. To this end, Khosla and coworkers systematically characterized the bottlenecks hindering successful AT swaps.[42] By analyzing the kinetics of AT acylation and KS-mediated condensation in native and engineered systems, they concluded that condensation in the hybrid modules occurred at a rate over ten times slower than the native system. In addition, limited proteolysis experiments and the poor expression of

the engineered proteins suggested that the AT swapped mutants are inherently less stable and adopt a different protein conformation. AT domain swaps may minimize or interfere with the important protein-protein contacts required for condensation. Swapped domains may disturb the orientation of stabilizing residues[64] with respect to hydrogen bonding partners that help maintain the dimeric structure of the PKS or prevent the proper interaction of the ACP domain with the KS-AT linker.[65] Most recently, Keasling and coworkers performed a detailed kinetic analysis of AT-swapped modules utilizing different domain boundaries and showed that the optimized boundaries could be applied to construct functional swaps of various heterologous AT domains.[66] We hypothesize that more comprehensive biochemical and structural studies of AT domain swaps will reveal a less ambiguous set of design rules for AT replacements.

A final method for AT domain engineering involves the complementation of inactivated *cis*-AT PKSs with trans-acting domains. Although this review does not focus on engineering *trans*-AT PKS systems, we will note that the freestanding AT domains from these PKSs can successfully communicate with *cis*-AT PKS modules.[38] In *trans*-AT PKSs, the acyltransferase domain is not incorporated into the same polypeptide as the remaining catalytic domains of a typical PKS module. Instead, a freestanding *trans* AT typically transfers extender units to multiple ACP partners. The utilization of *trans*-ATs to generate polyketide diversity in place of the activity of a native domain has been demonstrated in several contexts. Typically, an AT null (AT^0) mutant is generated by converting the conserved active site serine to an alanine, destroying the active site nucleophile responsible for initial attack of the extender unit thioester. Khosla and coworkers originally showed that *trans*-AT domains such as DszsAT from the disorazole PKS system could be used to complement an AT^0 version of DEBS1 *in vitro*.[38] It was later found that the DszsAT could also be used to produce fluorinated polyketides through the transfer of fluoromalonyl-CoA to several different AT-deficient modules of DEBS, both *in vitro* and *in vivo*.[67] This technique has since been further extended to a bimodular system in which DszsAT was able to communicate with multiple ACPs at once, transferring fluoromalonyl-CoA to AT^0 versions of DEBS module 2 and 3 to produce a difluorinated TKL.[68] Other *trans* AT domains, such as the KirCII AT, have successfully been used to install propargyl and allyl groups into polyketides, albeit so far only within the native PKS system.[44]

AT domain engineering has been extensively studied, yet similar to LMs, no consensus set of engineering rules exist. It is well established that mutation of the active site serine in a conserved "GHSxG" motif to alanine abolishes domain activity. Changing domain specificity with point mutations has also been successful, but rational

engineering in this case is still unpredictable. We believe that techniques such as saturation mutagenesis of active site residues can be successful but could require large screens of many mutants. AT domain swapping is currently also a viable method for changing α-carbon substituents. Based on recent experiments, it appears that screening and optimization of various swaps has yielded a potential set of widely applicable domain swap boundaries for ATs.[66] It remains to be seen, however, if this method will be generalizable to all type I PKS systems given the potential for disrupted protein-protein interactions and gatekeeping from downstream processing.

In order to incorporate rare functional groups via AT domain engineering, the desired extender unit must be provided either endogenously (integration of a biosynthetic pathway into the chosen heterologous host) or exogenously (feeding of an appropriate precursor that can be converted to a malonyl-CoA derivative). We refer the reader to the subsequent sections on host engineering for more details. Within this realm, it is important to note that many unusual polyketide extender units are constructed on freestanding ACPs. Although AT domain mutagenesis may result in more promiscuous AT domains, in the case of ACP-linked extender units (e.g. methoxymalonyl-ACP,[69,70] hydroxymalonyl-ACP, aminomalonyl-ACP, etc.[71]), AT domain swaps may be a superior engineering method because proper protein-protein interactions between the extender unit-carrying ACP and AT domain can be maintained.[72]

It remains to be seen whether domain swapping or site-directed mutagenesis will prove the more productive AT engineering method in a general. However, we anticipate that the continued improvement in DNA synthesis and sequencing technologies will facilitate more large-scale systematic experiments that result in the improved activities of engineered enzymes.

2.3.3. Ketoreductase domains

Ketoreductases perform the NADPH-mediated reduction of newly-formed β-keto groups after condensation. They are both stereospecific and stereoselective, and they are also known to epimerize α-substituents (if present) prior to reduction.[73,74] KRs are characterized based on their stereochemical outcomes: A-type KRs generate L-configured alcohols and B-type KRs D-configured (**Figure 2-4**). KRs are further described based on the final orientation of any α-substituents. A1 and B1 KRs produce D-configured α-substituents, and A2/B2, L-configured.[8] Finally, C-type KRs are reductase-

incompetent but can retain epimerase activity.[75] Structures are available for each type of KR: A1,[76] A2,[74] B1,[77] B2,[78] and C2.[79]

A- and B-type KRs can be distinguished not only by function, but also through the presence of certain structural elements. The presence of an "LDD" motif ~57 residues before the catalytic tyrosine indicates a B-type KR (although frequently only the final "D" of this motif is strictly conserved) whereas it is absent in A-type KRs. Additionally, A-type KRs typically contain a tryptophan eight residues before the catalytic tyrosine. Structural insight has guided mutagenesis efforts, leading to the generation of A2-type[74] and nonstereospecific A-type KRs[76] derived from A1-type KRs. Additionally, a B2-type KR was converted to an A2-type through a single point mutant.[80] Not only can KR specificity be modified through site-directed mutagenesis, but activity can be completely abolished as well. Mutation of the catalytic tyrosine of the DEBS module 6 KR lead to production of the expected ketone, illustrating that KR inactivation is a way to generate new polyketides.[81]

In addition to site-directed methods for engineering KR domains, numerous experiments have illustrated the viability of full KR domain swaps to modulate the stereochemistry and oxidation state of a given polyketide. Preliminary experiments were performed in the model DEBS system. A swap of the inactive KR domain from module 3 into module 2 resulted in the production of a TKL with a keto group in place of the alcohol generated by the native system.[82] This preliminary result showed that the surrounding domains function in the presence of a non-cognate KR partner. Complete removal of KR domains and replacement with an AT-ACP linker, on the other hand, has produced completely inactive mutants.[83] As the generation of loss-of-function mutants can easily be achieved via site-directed mutagenesis, it is the preferable method for KR inactivation because it allows for the retention of native protein-protein interactions and folding. Construction of KR-swapped PKSs with novel functions have also been reported. Successful gain-of-function KR swaps were implemented in a three module DEBS system in which KR domains from the rapamycin modules 2 or 4 were swapped into DEBS module 2, producing TKLs with non-native hydroxyl stereochemistry.[84] In a similar set of experiments, researchers from the same group showed that KR swaps could also be performed by replacing the KR in DEBS module 2 with KR5 and KR6 of DEBS.[85] Interestingly, KR6 was nonfunctional in this context. Further applying these KR swap design principles, Ashley and coworkers subsequently used KR-swapped DEBS modules to produce various derivatives of 6-deoxyerythronolide B *in vivo*, albeit with significant titer losses compared to the wild type system.[83]

More recent efforts have focused on applying KR swaps to tailor the stereochemistry of both hydroxy groups and the adjacent α-methyl groups. Weissman and coworkers evaluated 14 total A2 and B2-type KR swaps into module 2 of DEBS1+TE using different swap sites.[86] Interestingly, there were obvious differences in activity of swaps performed using the various junctions. In addition, the A-type KR swaps were more functional than B-type swaps, suggesting that domain replacements that retain the native KR type might be more successful (DEBS module 2 contains an A1-type KR). Finally, all of the KR domains were sourced from modules with KS-AT-KR-ACP domain organization, so swaps of KR domains with neighboring reductive domains may need to be optimized separately. A subsequent study also corroborated the results from the work by Weissman and coworkers. Only A-type KR domain exchanges into a lipomycin PKS derivative (LipPKS1+TE), natively containing an A2-type KR, were active whereas all attempts to swap in B-type KRs produced PKSs incapable of reduction.[87] By comparing KR swaps of domains that contain or lack the newly-discovered KR dimerization element (DE), Keasling and coworkers also found that retaining the DE from the acceptor module almost always produced PKSs competent at performing condensation. Retaining the DE produced mutants with much higher activity than previous reports, suggesting that more systematic analyses of the structural implications of domain swaps might improve the activity of hybrid PKSs.

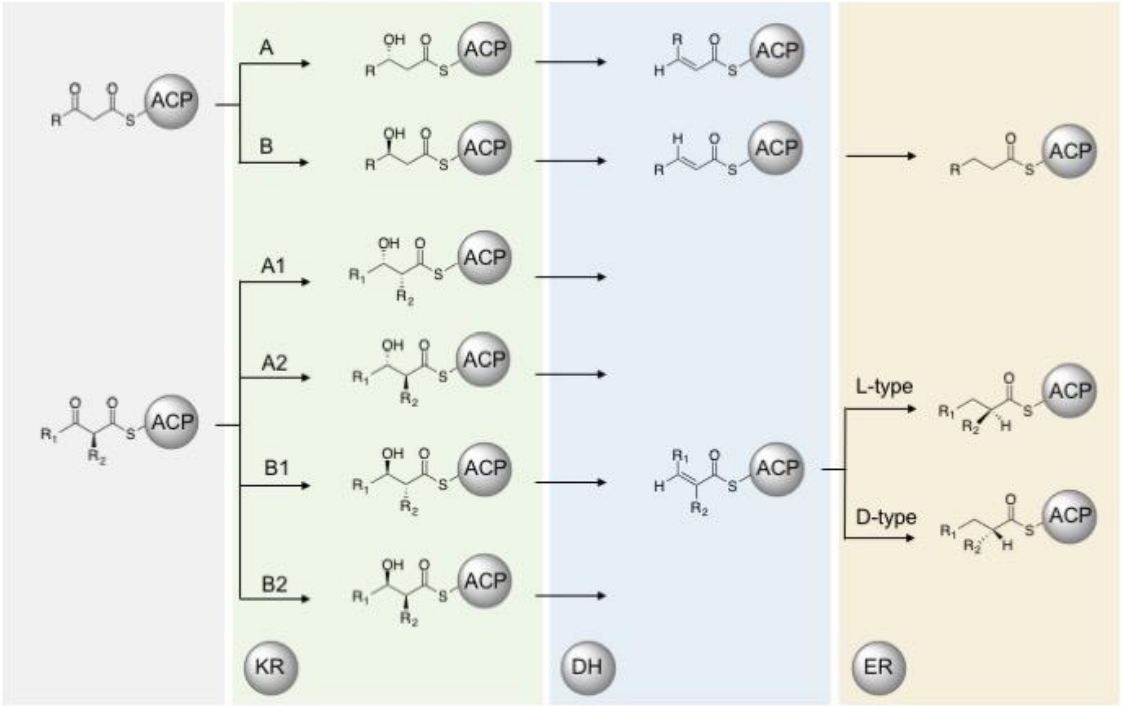


Figure 2-4 Potential product outcomes from PKS reducing domains. Potential stereochemical outcomes of each combination of β-carbon processing domains within a PKS module.

A few trends emerge when comparing the various strategies for altering the stereochemical outcome conferred by the KR domain. An aspect that appears to be generally consistent between studies with small molecule substrates, mutagenesis, and domain swaps is that KRs that reduce longer substrates (presumably possessing larger binding pockets) tend to be either less active on smaller substrates or tend to generate predominantly the energetically favored A2 product. This was observed in studies by Keatinge-Clay and coworkers, wherein small N-acetylcysteamine (SNAC) mimics were treated with a panel of various type I KRs.[74] This is also consistent with the work of Keasling and coworkers, where DEBS KR6 (the KR from the terminal module) was shown to retain the least amount of selectivity among the KR+DEs swapped.[87] The notion that expanding the binding pocket affects stereochemical outcome was further corroborated by the work of Keatinge-Clay via enlarging active site through point mutations.[80] This observation has implications for choice of donor KRs to exchange in KR swapping experiments. If the donor KR does not natively reduce a substrate of similar size to the target small molecule, it is unlikely to retain stereochemical fidelity, thus impacting the processing of downstream β-carbon processing domains (**Figure 2-4**).

2.3.4. Dehydratase and enoylreductase domains

Dehydratase domains are responsible for installing the majority of alkenes in polyketides. Dehydration proceeds via the *syn*-coplanar elimination of water and is therefore sensitive to the stereochemical configuration of the substrate.[88] Thus, DH domains are tied to the KR domains that precede them. Most DHs follow B-type KRs and catalyze *trans*-olefin formation (**Figure 2-4**).[8] However, both *cis*- and *trans*- alkenes are found in polyketide backbones. Post-PKS processing is implicated in formation of the *cis*-olefins in borrelidin[89] and rifamycin.[90] The *cis*-alkene of phoslactomycin, however, is likely installed by the DH of the first module of the synthase, Plm1.[91] This DH succeeds an A-type KR, so *syn*-coplanar elimination of water from the resultant KR product would yield a *cis* alkene (**Figure 2-4**). Despite having several DH crystal structures,[88,90,92] there is no clear trend in the active site residues that govern the stereochemistry of reduction or the preference for α-substituents.[93]

Although DH engineering has not been as extensively pursued as the engineering of other PKS domains, several examples of both site-directed and swap-based engineering exist. There are several known examples of naturally-inactive DH domains, such as those observed in the amphotericin,[94] avermectin,[95] nystatin,[96] and

nanchangmycin[97] clusters. Each of the inactive domains exhibits a H→ R or H→ Y mutation within the conserved active site motif "HxxxGxxxxPP". Installing the H→ Y mutation in a DH domain within the FR-008 PKS resulted in successful inactivation of the DH and production of solely the appropriately hydroxylated products.[98] Using this same technique, a DH in the nystatin cluster could also be inactivated.[99] DH swaps have also been attempted. The DH domain from DEBS was swapped into a module of the avermectin synthase in *S. avermitilis*, resulting in the exclusive production of C22,23-unsaturated avermectins (although at much lower titers than the wild type).[100] DH domain swaps were also employed in a polyketide synthase engineered to produce adipic acid, decreasing the amount of 3-hydroxyadipoyl product that had built up due to suspected non-optimal DH activity. Interestingly, the same group showed that DH domains can also be provided *in trans* to achieve the same effect (**Figure 2-1C**).[101]

Even less is known about the ER domains, which reduce *trans*-α,β-unsaturated intermediates provided by DHs. ERs, like KR domains, exhibit epimerase activity on α-substituents as well.[93] ERs that produce L-oriented products possess a conserved tyrosine that is absent in D-type KRs.[102] A crystal structure of the ER from the second module of the spinosyn PKS (Spn2) shows this characteristic L-type tyrosine residue proximal to 4-pro-*R*-hydride of the bound $NADP^+$, suggesting its role as the proton donor.[103] This structure also revealed a lysine-aspartate pair that was crucial for catalysis; the lysine is thought to act as the proton donor in D-type ERs. Finally, this structure revealed that the ER domain is inserted between the structural and catalytic domains of the KR, tying the ER to its cognate KR partner.

Early mutagenesis studies of ER domains in DEBS module 4 targeted the conserved "HAAAGGVGMA" NADPH binding motif for engineering.[104] Changing this sequence to "HAAASPVGMA," based on the NADPH binding motif of the inactive KR domain from the same module, resulted in production of primarily $\Delta^{6,7}$-anhydroerythromycin C, the expected product. Moreover, mutation of the conserved L-type tyrosine in the same ER resulted in a change of stereochemistry at the α-methyl substituent, switching it from the S to R configuration.[102] Nevertheless, it should be noted that the same mutation in the rapamycin ER failed to produce the expected inversion of stereochemistry in TKL products; therefore, more basic studies of stereocontrol in ERs are needed before a comprehensive method for switching α-substituent orientation can be determined. Subsequent mutagenesis experiments performed by Keatinge-Clay et al. were unable to completely inactivate the Spn2 ER through a single mutation.[103] Instead, mutations both slowed the rate of reduction and increased the rate of the hydroxylation side reaction.

When introducing a non-native DH and ER domain into a chimeric PKS, it is important to consider the DH stereoselectivity of the hydroxylated product exercised by the upstream KR domain (**Figure 2-4**). It might be best to pair DH domains with their cognate KR partners in order to ensure proper stereochemistry of the hydroxylated substrate. Alternatively, swaps of DH domains into modules with the same KR type could also be successful. Not only is proper substrate stereochemistry required, but substituents farther from the zone of reactivity could also come into play.[101] Another study corroborated this hypothesis, presenting structural and functional data to suggest that DHs which natively act on long-chain (C28-30) substrates are unable to dehydrate shorter chain (C4) substrates.[105] This was consistent with the structural information presented, highlighting a substrate channel that made hydrophobic interactions with the acyl chain of the long-chain substrates. Other issues involving protein-protein interactions could also arise, as different ACP partners can completely reverse the stereospecificity of DHs.[90]

2.3.5. Reductive loops

Reductive loops, here defined as any combination of reducing domains, determine the degree of reduction of the β-carbon formed post-condensation and also typically set the stereochemistry of any α-substituents (see previous sections). Typically, site-directed mutagenesis is performed on single reductive domains to achieve a desired outcome, whether it be domain inactivation or a change in activity. Schulz and coworkers, however, systematically evaluated the feasibility of "domain skipping" mutations in every reducing domain in the monensin PKS in order to evaluate the ability of downstream modules to accept less reduced substrates.[106] PKSs mutated in a way that produced less reduced intermediates were less likely to produce the desired final product. All ER inactivations reliably produced the expected products, but only two thirds of DH and one half of KR inactivations resulted in any detectable premonensin analogs. This suggests not only that downstream domains may display selectivity for their native substrates, but also that attempts to engineer more drastic changes of degree of reduction within the molecular structure may be less successful.

Full reductive loop additions into modules with fewer (or no) native reducing domains have been successfully employed in a few cases. A series of reductive loop swaps into DEBS module 2 were promising examples, demonstrating that heterologous reducing domains can be feasibly employed.[107] Excitingly, Leadlay and coworkers illustrated that the KR, DH and ER domains can not only be exchanged among PKSs, but

they can also be added to modules having a partial reductive loop, natively containing a KR domain. However, the junctions of each domain swap may need to optimized on a case-by-case basis, as no single set of boundaries was optimal for every hybrid PKS. By heeding this advice, Keasling and coworkers were successfully able to engineer a PKS capable of producing a novel polyketide, adipic acid, by introducing a full reductive loop into the first module of the borrelidin PKS (**Figure 2-1B-C**). A series of donor loops and junction sites were tested to find the most productive mutant. Although application of reductive loop swaps was promising, a significant bottleneck at the DH was observed.[101]

In addition to full reductive loop swaps, other examples of reductive domain replacements have been published. A swap of the DH-KR didomain from module 11 to module 12 of the pimaricin PKS improved activity of the only slightly active module 11 DH.[108] The most successful mutant retained the linkers flanking the native DH-KR, while replacing simply the DH domain proved ineffective. Additional examples include a swap of the pikromycin module 4 DH-ER didomain into the avermectin PKS[109] and replacement of the DEBS KR domains with the full reductive loop from rapamycin module 1,[83] both of which resulted in significant decreases in product yield *in vivo*. Since having a match between substrate size and activities of both the KRs and DHs (*vide supra*) have been shown to impact their levels of activity on non-native substrates (the ER has been less extensively studied), it may be important to introduce reducing loops that act on native substrates of a similar size with regard to the target donor molecule. By continuing to parse the limitations of these types of swaps, we imagine that more successful implementations of reductive domain additions and replacements will arise.

2.3.6. Offloading domains

To complete PKS biosynthesis, a terminal domain must release the ACP-bound polyketide intermediate in the last module. Termination domains include thioesterase (TE) domains and reductase (R) domains (**Figure 2-2, 2-5**). The more common of the two release mechanisms, TEs, belong to the α/β-hydrolase superfamily of proteins and can release the polyketide as a linear acid, macrolactone, macrolactam, or macrothiolactone[9] (**Figure 2-2, 2-5**). They are generally selective catalysts in the presence of native substrates, exemplified by the highly regio-[83] and stereospecific[110] cyclization performed by DEBS TE. Despite their native specificity, cyclization TEs are often promiscuous towards unnatural substrates.[12] DEBS TE, for example, cyclizes various macrolactones derived from engineered versions of DEBS that produce non-native oxidized and substituted backbones.[83] Additionally, non-native 6-, 8-, 12-, 14- and 16-membered rings have been produced by either DEBS TE fusions or mutasynthesis.[85,111,112] Cyclizing

TEs (especially the commonly used DEBS TE) are also known to hydrolyze nonnative substrates in engineered systems when cyclization is not feasible.[12,21,101,113,114]

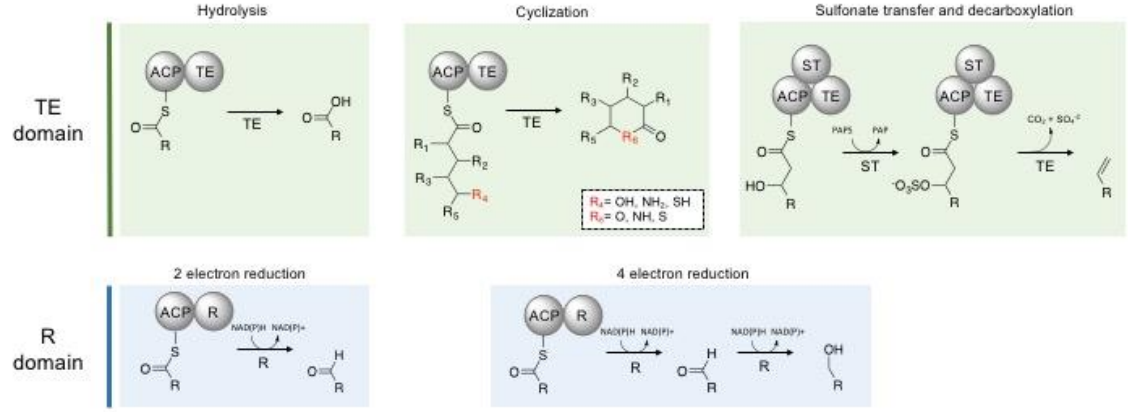


Figure 2-5 Offloading domains. Common TE-mediated release include products such as linear acids, lactones,lactams, thiolactones and olefins. The less common R-domains conduct a two-electron reduction to produce aldehyde final products or primary alcohols through a four electron reduction.

A number of striking differences emerge between the TEs that natively hydrolyze their substrates and those that cyclize. The best studied of the hydrolytic TEs is from tautomycetin, which has been structurally characterized and shown to have a narrower substrate chamber than that observed in the structures of the cyclizing pikromycin and DEBS TEs.[115] The selectivity of this TE was further elucidated with chemically-synthesized SNAC intermediates, which demonstrated that its hydrolytic activity is highly stereospecific for *R*-β-hydroxy moieties. Further probing the hydrolytic selectivity of the tautomycetin TE, Kim and coworkers swapped the tautomycetin TE domain with the macrocyclic polyketide pikromycin TE.[116] The Pik-TE-swapped strain produced a mixture of both linear tautomycetin and a cyclized analog.

A termination mechanism that is evolutionarily related to a thioesterase consists of tandem sulfotransferase (ST)-TE domains which yield terminal olefins, as is found in the curacin PKS (**Figure 2-2, 2-5**). The ST-TE didomains mechanistically employ the sulfotransferase domain to activate the β-hydroxy group formed by the terminal module. The activated sulfate group then undergoes a decarboxylative elimination reaction to yield the terminal olefin as the released product.[92,117,118] This biosynthetic logic is particularly attractive for applications of engineered polyketides to produce a variety of desirable chemicals, especially monomers for polymeric materials[119] or synthetic handles for further functionalization of natural products.[120]
The less common of the two releasing mechanisms, R domains, generally catalyze the NAD(P)H-dependent reductive release of polyketide and non-ribosomal peptide

intermediates via thioester reduction (**Figure 2-2, 2-5**). R domains possess a Rossmann fold,[121] characteristic of nucleotide binding proteins which release acyl intermediates by either 1) a 2- electron reduction, yielding an aldehyde final product or by 2) a 4-electron reduction, yielding a primary alcohol final product. Examples of these include the reductive release of aldehyde-containing product aureusimine[122] and primary alcohols in glycopeptidolipid[123] and myxalamid biosynthesis.[9,124] Structure-guided mutagenesis of the terminal myxalamid R domain, MxaA, generated variants with increased activity towards hydrocarbon substrates that resulted in highly-reduced primary alcohols.[125] The removal of bulky residues in the C-terminal substrate binding domain facilitated reduction of non-native 10-carbon intermediates, yielding a reduced 10-carbon alcohol product.[125] This termination mechanism is of particular interest when trying to engineer polyketides that generate fuel molecules (e.g. n-butanol[126]).

2.3.7. Intermodular linkers

Intermodular linkers facilitate communication and chain translocation between two modules. They can take the form of C- and N-terminal docking helices that mediate contacts between modules on separate polypeptides, or they can consist of simple peptide linkers between two modules on the same polypeptide (**Figure 2-2**).[10,127–129] Early attempts to improve intermodular communication between PKSs focused on covalently fusing two non-covalently-linked modules together. The expected TKLs could be produced in *S. coelicolor* by fusing DEBS module 1 to DEBS modules 3 or 6 by conserving the natural intermodular linker between modules 1 and 2. This technique could also facilitate communication of module 1 with rifamycin module 5 (Rif M5). Finally, by appending the C-terminal linker domain from DEBS module 2 to the end of Rif M5, the hybrid DEBS M1-Rif M5 protein could be coaxed to communicate with the downstream DEBS M3. This preliminary work laid the foundation for combinatorial biosynthesis of polyketides utilizing modules, not domains, as the fundamental unit.[130]

Khosla and coworkers demonstrated that the position of modules within their native context is an important consideration for intermodular linker engineering. Swapping linker regions between module 2 and module 3 of DEBS with the module 4 and 5 linkers significantly increased the K_m of the module-module interaction with little to no effect on the k_{cat} of polyketide transfer between the modules.[131] The Khosla group also suggested that the KS domains of C-terminal modules, such as DEBS module 2 and 4, are less selective of substrates from upstream ACP partners than N-terminal modules because C-terminal KSs are already covalently tethered to the upstream ACP. The importance of ACP-KS interactions for chain transfer were again highlighted in the work

of Santi and coworkers, where over 150 module-module interactions were tested for chain transfer.[132] The efficacy of 11 upstream and 14 downstream module combinations, facilitated by the C- and N-terminal linker domains from DEBS module 2 and 3, respectively, were tested for production of the expected TKL products. Although most of the individual PKSs were active on some level, less than 50% of the combined two module systems produced any product. The best combinations were between native partners, suggesting that incorrect ACP-KS interactions were a limiting factor.[133]

2.4. Engineering at the host/cellular level

2.4.1. Precusor-directed biosynthesis and mutasynthesis

Despite all of the current insights gained through PKS engineering at the amino acid, domain, and module level, progress has been slower than expected. This arises from a lack of understanding and inability to predict the tolerance of PKS enzymes to utilize non-native precursors as building blocks. Non-native precursor incorporation requires a deep understanding of the gate-keeping activities of PKS enzymatic machinery. Precursor-directed biosynthesis (PDB) has been a successful tool utilized to further understand and expand PKS substrate promiscuity by feeding analogues of their natural building blocks that are likely tolerated by the native biosynthetic PKS in the producing host. The efficiency of PDB can be enhanced by complementation with mutational biosynthesis ("mutasynthesis"), wherein the naturally occurring precursor pathways are inactivated, thus removing competition from natural precursors. In addition to precursor supplementation through feeding experiments, metabolic pathways that produce precursor analogs can be introduced heterologously to replace the deleted pathways.[134]

One of the earliest and most successful examples of mutasynthesis targeted the avermectin PKS. A strain of *Streptomyces avermitilis* was generated wherein the enzymes required for generating the precursors 2-methylbutyryl-CoA and isobutyryl-CoA were inactivated, more specifically the branched chain fatty acid dehydrogenase complex *bkd*. Out of more than 800 potential precursors tested, over 40 starter unit analogs were tolerated by the avermectin PKS.[20] Via this method, a cyclohexyl-containing avermectin derivative (later named doramectin) was generated that exhibited increased antiparasitic activity against veterinary pathogens.[135] Another group later identified a shikimate-derived cyclohexyl-CoA biosynthetic pathway that, when re-introduced into *S. avermitilis* Δ*bkd*, enabled the production of doramectin without cyclohexanoic acid supplementation.[136,137] Intriguingly, while other PKSs are primed with isobutyryl or 2-methylbutyryl-CoA (such as lipomycin and tautomycin), no analogous

experiments to generate other mutasynthetic analogs in *Δbkd* strains has been reported. While the lipomycin loading AT has been shown to be somewhat promiscuous, only six loading substrates have been characterized.[21] Thus, it is unclear if the avermectin PKS has an unusually promiscuous loading AT, or if this extreme promiscuity is a general feature of loading ATs from type B LMs that accept bulkier acyl-CoA priming substrates.

Other examples of PDB have targeted the macrolides rapamycin[138,139] In rapamycin biosynthesis, Leadlay and coworkers conducted feeding experiments with starter-unit analogues. A series of 21 substrates consisting of monocyclic, polycyclic and branched aliphatic acids, substituted benzoic acids and heterocyclic acids, were fed to the rapamycin-producing strain of *S. hygroscopicus*.[139] New metabolites were observed in the presence of three monocyclic aliphatic acids, demonstrating a certain degree of starter-unit promiscuity within the rapamycin LM and further suggesting some substrate tolerance within the downstream modules (**Figure 2-6A**).[139] In a similar approach, Martin and coworkers explored carboxylic acid starter-unit tolerance in the spinosyn PKS appended with the avermectin or erythromycin LM.[140] Supplementation with a range of carboxylic acids led to the successful production of new spinosyn analogues (**Figure 2-6B**).[140] In a similar example, the aureothin PKS pathway demonstrated starter unit tolerance towards various aromatic priming analogs.[138] Lastly, Khosla and coworkers evolved an engineered erythromycin biosynthetic pathway in *E. coli*, yielding erythromycin analogs from SNAC precursors. Khosla and coworkers fed diketide SNAC precursors to *E. coli* cells harboring a truncated form of DEBS lacking the loading and first modules. They demonstrated the incorporation of the native substrate analog diketide SNAC and a novel diketide-SNAC containing an alkynyl moiety into the erythromycin final product (**Figure 2-6C**). These results support the tolerance of DEBS to utilize a nonnative substrate for priming, extension and modification by the downstream DEBS PKS machinery to produce an alkynyl-containing erythromycin derivative.[141,142]

The selectivity and promiscuity of extending acyltransferase domains can also be targeted for PDB and/or mutasynthetic approaches to alter PKS scaffolds. While the AT domains that select for malonyl-CoA or methylmalonyl-CoA are typically quite selective for their respective extender units,[40] it is common for ATs that naturally select for more uncommon and/or bulkier substrates to have more relaxed intrinsic substrate selectivity. For example, module 5 of the monensin PKS natively incorporates both methylmalonyl-CoA as well as ethylmalonyl-CoA.[143–145] In the salinosporamide PKS, while the major product is derived from SalA-AT-mediated chloroethylmalonyl-CoA incorporation (forming salinosporamide A), analogs containing ethylmalonyl-CoA (salinosporamide B) and methylmalonyl-CoA (salinosporamide D) are also generated (**Figure 2-6D**).[146] This

inherent promiscuity can be harnessed in concert with precursor pool engineering to bias the distribution of metabolites or generate new metabolites altogether. Deletion of 5-chlorodeoxyadenosine (a precursor for chloroethylmalonyl-CoA) biosynthetic genes coupled with exogenous supplementation of 5-fluorodeoxyadenosine[147] or heterologous expression of the heterologous precursor pathways[148] resulted in the generation of a fluorinated analog, fluorosalinosporamide .

Likewise, the promiscuity of the monensin AT5 was explored to generate premonensin analogs (**Figure 2-6E**). Schulz and coworkers determined that MonAT5 accepts propargylmalonyl, allylmalonyl, propylmalonyl, and butylmalonyl-SNAC when supplemented in the media.[145] Similar promiscuity was shown for an ethylmalonyl-CoA selective AT in the kirromycin pathway, KirCII (**Figure 2-6F**). *In vitro*, KirCII accepts a range of longer malonyl derivatives including allylmalonyl-CoA, propargylmalonyl-CoA, and to a lesser extent azidoethyl-CoA.[149,150] Rather than supplementing the culture with SNAC analogs, Williams and coworkers introduced a highly promiscuous mutant of the malonyl-CoA ligase, MatB, from *Rhizobium trifolii*[150,151] into the host strain *S. collinus* Tü 365 and supplemented exogenous allylmalonate and propargylmalonate. The malonate derivatives were then activated by MatB and incorporated into the polyketide to generate kirromycin analogs.[44] Propargyl-containing extender units were introduced in one additional polyketide scaffold through targeted mutagenesis. A combination of structural modeling and sequence comparisons of ATs with specificity for bulkier substrates (such as ethylmalonyl-CoA) was used to expand the promiscuity of a methylmalonyl-selective AT.[53,56]

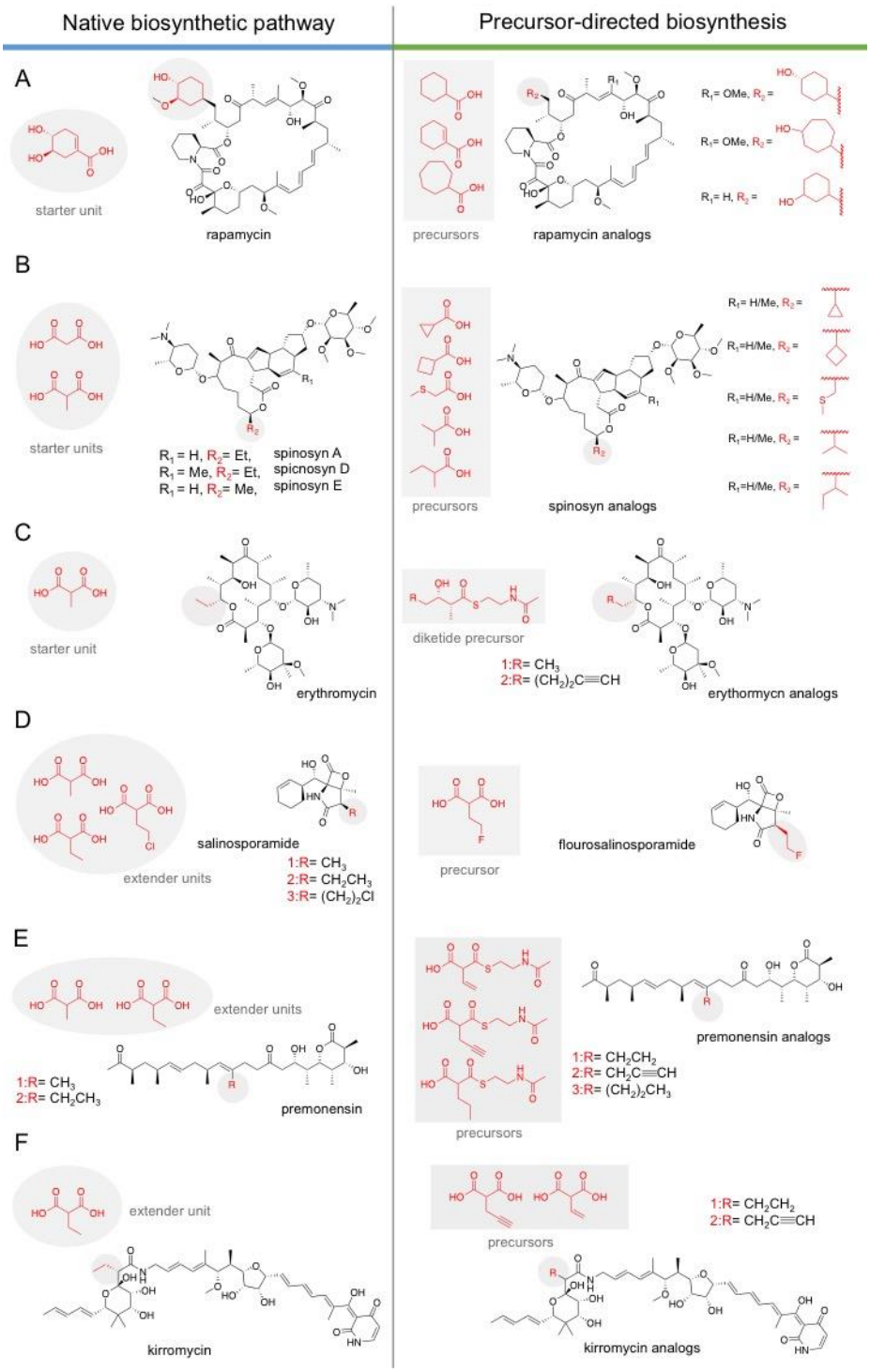


Figure 2-6 Examples of precursor-directed biosynthesis. A) Starter unit PDB in the rapamycin PKS. **B)** Starter unit tolerance in spinosyn biosynthesis. **C)** Extender unit PDB using a diketide SNAC precursor in the erythromycin PKS. **D)** Incorporation of fluoromalonyl-CoA through mutasynthesis of the salinosporamide pathway. **E)** Incorporation of propargyl-, propyl- and allylmalonyl-SNAC into the monensin polyketide backbone. **F)** Incorporation of propargyl- and allylmalonyl-CoA into the kirromycin polyketide backbone.

2.4.2. Metabolic engineering for improved precursor pools

Outside the limitations of directly engineering of PKSs, one approach to improving polyketide production involves increasing the availability of polyketide precursors. Pathways for the biosynthesis of polyketide metabolites often use precursors synthesized during glycolysis, the tricarboxylic acid cycle, and the pentose-phosphate pathway.[152] Metabolic engineering has successfully improved selected bottlenecks in various strains, increasing intracellular precursor pools, and thus redirecting flux towards the desired polyketide biosynthetic pathway. The deletion of the *S. coelicolor* phosphofructokinase gene *plfkA* or the *S. lividans* glucose-6-phosphate dehydrogenase genes *zwf1* and *awf2* achieved higher production of actinorhodin and undecylprodigiosin, respectively.[153] Inactivation of the glyceraldehyde-3-phosphate dehydrogenase *gap1* in *S. clavuligerus* resulted in more clavulanic acid production.[154,155] In contrast to deletion or knockout of genes, overexpression and knockin of genes has been utilized more often as a strategy to increase specific precursors for polyketide production. Such polyketide precursors include well-known and unique starter-/extender-units. For many type I assembly-line PKS pathways, such as DEBS, priming begins with one propionyl-CoA starter unit and extension is performed with (2*S*)-methylmalonyl-CoA. *E. coli* does not naturally produce (2*S*)-methylmalonyl-CoA, posing a bottleneck for expression of the erythromycin pathway in this host.[156] Hillenmeyer and coworkers have worked to overcome *E. coli*'s (2S)-methylmalonyl-CoA limitation by introducing heterologous propionyl-CoA carboxylases (PCCs) (**Figure 2-7A**).[157] The propionyl-CoA carboxylase is a biotin-dependent enzyme that catalyzes the carboxylation of propionyl-CoA to form (2S)-methylmalonyl-CoA. The core PCC complex consist of two genes encoding the biotin carboxyl carrier α-subunit protein and a β-subunit containing the carboxylase activity.[158,159] Hillenmeyer and coworkers examined the effect of heterologous expression of various PCCs on erythromycin production in *E. coli* by optimizing both expression of the *S. coelicolor* PCC complex and screening 13 homologous PCCs from diverse species.[157] The hybrid PCC complex from *Myxococcus*

fulvus and *S. coelicolor* outperformed all other PCC complexes in stimulating erythromycin production.

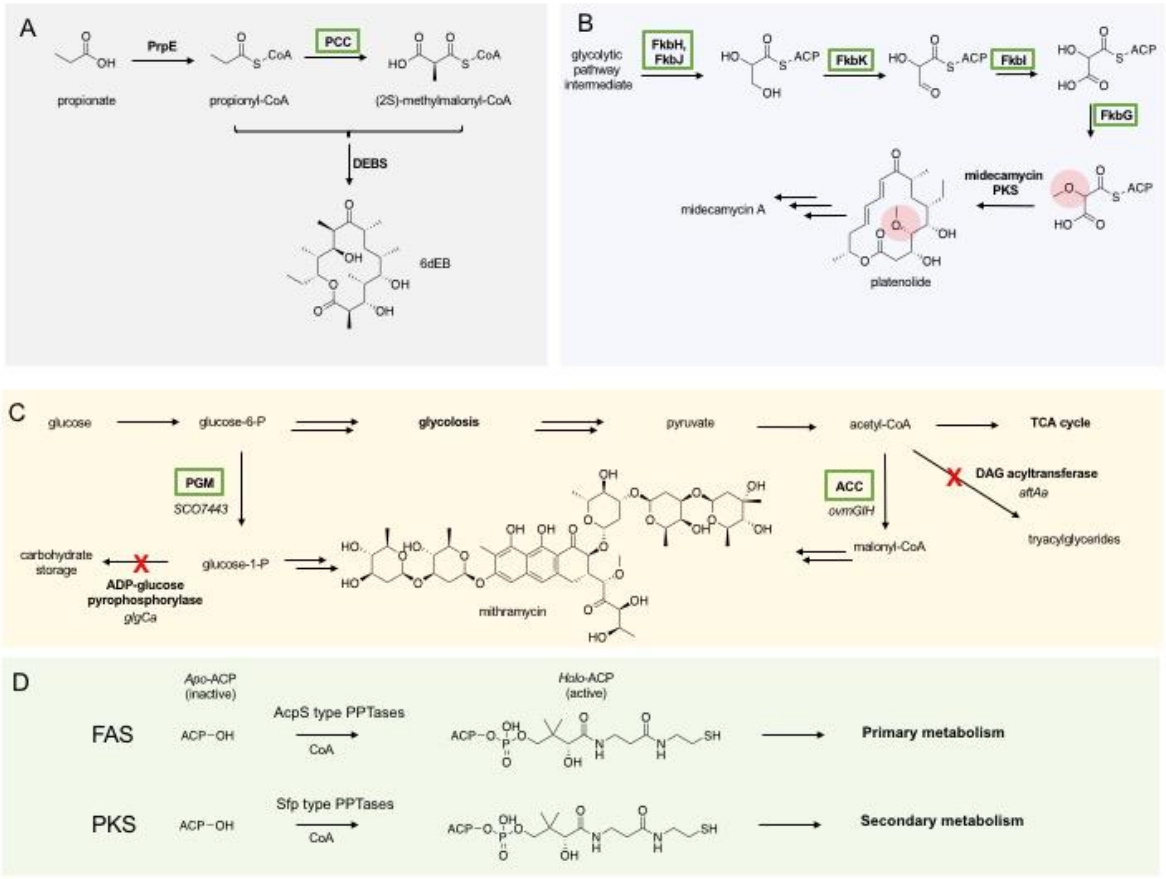


Figure 2-7 Examples of metabolic engineering for improved precursor pools. A) Overexpression of PCC for the conversion of propionyl-CoA to (2S)-methylmalonyl-CoA allowed for the production of 6-deoxyerythronolide B in bacteria that don't natively produce methylmalonyl-CoA (such as *E. coli*). B) Expression of the methoxymalonyl-ACP biosynthetic pathway in a platenolide producer generated a platenolide analog containing the methoxy moiety. C) Higher titers of mithramycin were achieved by increasing the precursor supply of malonyl-CoA and glucose-1-phosphate. Green boxes denote upregulated pathways and red cross marks denote downregulation or knockout pathways. D) Post-translational modification of *apo*-ACPs by PPTases generated active *holo*-ACPs which can activate FAS biosynthetic pathways in primary metabolism or PKS pathways in secondary metabolism.

Unlike *E. coli*, where certain acyl-CoA extender units may be limited or nonexistent, *Streptomyces* already contain the common precursors to support polyketide biosynthesis. The diverse precursor pools that *Streptomyces* produce facilitate the heterologous expression and engineering of full pathways in these hosts. For example, in *S. coelicolor* and *S. lividans* strains heterologously expressing the

DEBS PKS, Floss and coworkers constructed strains capable of producing 3 mg/L of 2-desmethyl-2-methoxy-6-deoxyerythronolide via the introduction of the methoxymalonyl-ACP biosynthetic pathway from *Actinosynnema pretiosum*.[63] The incorporation of methoxymalonyl-ACP was also utilized in *S. fradiae*, a host that natively produces three of the most common polyketide precursors: malonyl-CoA, methylmalonyl-CoA, and ethylmalonyl-CoA.[160] By introducing the biosynthetic pathway for methoxymalonyl-ACP into an *S. fradiae* strain heterologously expressing the midecamycin pathway, Katz and coworkers produced a methoxymalonate-platenolide analog (**Figure 2-7B**).[160] More recently, precursor pool engineering was also successfully employed to improve salinomycin production in *S. albus*. Metabolomic analysis suggested that intracellular ethylmalonyl-CoA concentrations were limiting titer, so an additional copy of the crotonyl-CoA carboxylase gene responsible for generating ethylmalonyl-CoA was integrated into the producing strain. This approach improved titers over tenfold.[161]

Méndez and coworkers improved precursor metabolite pools for the production of the antitumor compound/polyketide mythramycin in *S. argillaceus* by increasing the precursor supply of malonyl-CoA and glucose-1-phosphate (**Figure 2-7C**).[155] The mythramycin PKS uses ten malonyl-CoA units for chain extension and cyclizes the extended polyketide to a tetracyclic intermediate. The tetracyclic intermediate is subsequently glycosylated by five deoxysugars derived from glucose-1-phosphate, generating mythramycin as the final product.[155] By either overexpressing the *S. coelicolor* phosphoglucomutase gene *pgm* or by inactivating the ADP-glucose pyrophosphorylase gene *glgCa,* they were able to increase glucose-1-phosphate production (**Figure 2-7C**). Moreover, they were able to increase malonyl-CoA by overexpressing the acetyl-CoA carboxylase *ovmGIH* gene or by inactivating the acyl-CoA diacylglycerol acyltransferase gene *aftAa*, the latter strategy being superior for improving mythramycin production.[155]

2.4.3. Phosphopantetheinyl transferase expression and regulation

Another frequent issue encountered when heterologously expressing PKSs is the need for post-translational activation of each ACP domain with a prosthetic phosphopantetheine arm. All ACP domains must be converted to the *holo* form by phosphopantetheinyltransferases (PPTases), which install the phosphopantetheinyl prosthetic group of ACP domains. PPTases can have a range of substrate selectivities, depending on whether they evolved to activate a carrier protein from a specific biosynthetic pathway or whether they have a pleiotropic role in the cell, activating multiple pathways.[162–164] PPTases are classified by whether they activate fatty acid

biosynthetic pathways from primary metabolism (AcpS-type PPTases) or polyketide and nonribosomal peptides from secondary metabolism (Sfp-type PPTases) (**Figure 2-7D**).[165] In host organisms that do not natively harbor multiple polyketide and/or non-ribosomal peptide biosynthetic pathways (such as *E. coli* or *S. cerevisiae*) integration of a promiscuous Sfp-type PPTase is critical to enable the heterologous production of polyketide metabolites. The promiscuous PPTase Sfp, from the surfactin biosynthetic pathway in *Bacillus subtilis*, has historically been the most common choice for application in heterologous systems.[166-167] Sfp has been integrated into *E. coli* to generate strains engineered specifically for polyketide production: BAP1[168] and K207-3.[169] Successful production of polyketide metabolites has also been accomplished in *S. cerevisiae* strains containing genomically-integrated Sfp.[170–172] The usefulness of many *B. subtilis* laboratory strains for heterologous PKS production is diminished because of a common frameshift within the *sfp* gene, effectively shutting down the production of multiple secondary metabolites.[163,173,174] Indeed, to successfully heterologously produce 6-dEB in the commonly used *B. subtilis* Marburg 168 strain, restoration of Sfp activity was required.[174] As an alternative to Sfp, the similarly promiscuous PPTase, Svp, from *S. verticullus* ATCC15003 has also been used to post-translationally modify a wide range of ACPs.[175]

In most *Streptomyces*, heterologous polyketide production can be achieved without integration of an exogenous PPTase,[176] suggesting that most *Streptomyces* species natively harbor promiscuous PPTases. This is unsurprising, given the wealth of secondary metabolites that are produced by Actinomycetes. However, more recently the overexpression of PPTases has been explored as a means of boosting polyketide production in various Actinomycetes. Li and coworkers demonstrated that overexpression of the endogenous PPTase SchPPT in *S. chattanoogensis* L10 increased the production of the polyketide natamycin by over 40%.[177] Other experiments suggest that PPTases may play a major role in the regulation of PKS expression. For example, overexpression of either Sfp or Svp in over 33 Actinomycetes activated the expression of silent biosynthetic pathways.[178] Although a basal level of PPTase expression is typically sufficient for polyketide production, these experiments suggest that overexpression of these enzymes can significantly enhance or alter the regulation and production PKS metabolites. Thus, to achieve extremely high titers of polyketide products in *Streptomyces* and related Actinomycetes, manipulating the expression of both endogenous and exogenous PPTases is likely an underexplored approach.

2.4.4. Transcriptional regulation and refactoring of PKS genetic components

Polyketide biosynthetic gene clusters commonly contain pathway-specific regulatory elements that act as activators or repressors of various genetic elements within the cluster. In the biosynthesis of tylosin, for example, both activators and repressors have been identified and characterized.[179,180] Activator elements often belong to the *Streptomyces* Antibiotic Regulatory Protein (SARP) family, a group of proteins known to enhance the production of polyketides such as daunorubicin, mithramycin and tylosin.[153] These SARPs can be effectively used as tools to increase polyketide production. Tylosin production, for example, is increased when the SARP regulators *tylS* and *tylR* are overexpressed in the native producer *S. fradiae*.[180] Mithramycin production in *S. argillaceus* can also be enhanced by overexpressing two SARPs, *mtrY* and *mtmR*, on multicopy plasmids.[181,182] Other families of transcriptional regulators, such as the LysR-type transcriptional regulators (LTTRs), can serve as additional tools for controlling polyketide production. The role of LTTRs has been studied in the ascomycin biosynthetic cluster; in particular, Wen and coworkers have characterized the role of the LTTR *fkbR1*.[183] Inactivation of *fkbR1* leads to a reduction of ascomycin production, which can be restored by complementation with *fkbR1*. Overexpression of *fkbR1*, on the other hand, results in increased ascomycin production, suggesting that *fkbR1* acts as a positive regulator.[183] Manipulation of negative regulators can also be used as a tool for manipulating polyketide pathways. For example, inactivation of a transcriptional repressor (*tylQ*) from the tylosin pathway causes production to begin at an earlier stage of *S. fradiae* growth.[184] On the other hand, disruption of the *tylP* gene that encodes a γ-butyrolactone receptor increases production.[179]

The use of DNA-based regulatory elements can be used to complement the manipulation of protein-based positive and negative regulators for the optimization of polyketide production. The introduction of heterologous promoters, for example, has been effectively used to activate multiple silent polyketide clusters in various hosts. Eliciting expression of silent polyketide pathways is often challenging due to poorly understood regulation mechanisms. In addition, these silent polyketide gene clusters often remain inactive under typical culturing conditions. With the goal of developing a generalizable method to elucidate the function of cryptic biosynthetic gene clusters, Zhao and coworkers developed a synthetic biology approach to activate silent biosynthetic gene clusters.[185] The entire silent biosynthetic gene pathway was refactored using a plug-and-play scaffold, containing a set of heterologous functional promoters and placed in a heterologous host under more traditional culturing conditions. Using this strategy, Zhao and coworkers successfully activated the previously silent spectinabilin pathway from *S. orinoci*.[185]

More recently, Apel and coworkers used a similar plug-and-play method for the assembly of artificial genes into functional biosynthetic gene clusters.[186] As a proof of concept, the novobiocin pathway was disassembled and genetically reorganized using the artificial gene operon assembly method, a tool that consecutively assembles artificial genes into a destination vector and subsequently expresses them under the control of a de-repressed promoter in a *Streptomyces* host. By completely refactoring the pathway in *S. coelicolor*, Apel and coworkers were able to observe production of novobiocin precursors and novobiocin.[186]

The increasing promise of the CRISPR-Cas9 system for rapid and targeted genetic manipulations[187,188] has led to the development of protocols for using the technique in natural product-producing bacteria such as *Streptomyces*. Several groups laid the foundation for work with CRISPR-Cas9, first with the development of methods for performing knockouts.[189,190] Later, Zhao and coworkers extended the usefulness of this CRISPR-Cas9 platform, using it to perform knockins to activate silent biosynthetic gene clusters in *Streptomyces*.[191] In a one-step strategy, multiple biosynthetic gene clusters of type I, II and III PKSs in five *Streptomyces* species were activated. Moreover, by introducing strategically-placed constitutive promoters, they elicited the production of a new metabolite in *S. viridochromogenes*: a multicyclic type II polyketide.[191]

2.5. Host, precursor, and protein engineering synergy

Ultimately, for PKS protein engineering to be feasibly employed to produce target chemicals, it must be evaluated whether the engineered PKS activity *in vivo* is similar to its activity in model systems, such as biochemical *in vitro* assays. *In vitro* studies afford absolute control of substrate and cofactor concentrations, which is not feasible *in vivo*. Thus, combinations of host and engineered PKS can result in various (and sometimes unanticipated) substrate incorporations. A study that exemplifies this phenomenon dissects the LM of DEBS. While in the native *Saccharopolyspora erythrea* DEBS is primed exclusively by propionyl-CoA,[192] the LM of DEBS incorporates other priming units in heterologous systems.[193] Depending on the host, TKLs from DEBS1+TE can form via priming by acetyl-CoA (in *S. coelicolor*[194] and *S. venezuelae*[193]) or isobutryryl-CoA (in *S. venezuelae*), albeit as minor products. Reynolds and coworkers interrogated the relationship between PKS architecture and starter unit incorporation in variations of DEBS1+TE expressed in *S. venezuelae*. It was determined that placing the LM and module 1 on separate polypeptides (with the use of heterologous docking domains) changed the distributions of acetyl, propionyl, and isobutyryl priming units incorporated in comparison to the ratios observed with both modules on the same polypeptide (i.e. the selectivity shifted from predominantly propionyl selective to

isobutyryl-selective). This was rationalized through biochemical studies suggesting that the loading AT selects isobutyryl-CoA as the thermodynamic product. However, DEBS KS1 has faster kinetics for propionate compared to isobutyrate. When the loading domain is placed on a separate polypeptide from module 1, there is a resulting kinetic stall. This kinetic stall shifts the product formation towards thermodynamic control, especially when higher levels of isobutyl-CoA are present.[193] This example demonstrates that the apparent selectivity of PKS domains can be altered depending on the architecture of the PKS in concert with the precursor availability. Therefore, results of both *in vitro* and *in vivo* experiments for engineered polyketide production should be evaluated with a careful consideration of the advantages and limitations of each experimental approach.

2.6. Conclusions

With all the current success in PKS/host engineering, the gap between our understanding of PKS biosynthetic logic and the application of PKSs as a retrosynthetic platform is narrowing. In the past, our ability to design, build and test chimeric PKSs was limited by DNA sequencing and assembly techniques. Currently, our ability to discover and generate PKS genotypic diversity, enabled by DNA sequencing and synthesis technologies, allows us to rapidly build sizeable libraries of engineered PKSs.[195,196] However, our ability to screen the resulting libraries lags. There is a need for more high-throughput approaches to test engineered PKS scaffolds. Analytical methods such as liquid chromatography coupled with mass spectrometry are improving and new techniques are being developed which allow for increased throughput with limited sample preparation.[14,197–199] Transcription factor-based biosensors could be potentially employed to screen and select for successful engineered PKSs constructs.[200] Additionally, there are an increased number of computational and bioinformatic tools which aid the identification of candidate swaps as well as the curation and annotation of PKS pathways.[201–206] With the increasing success of chimeric PKSs scaffolds, there will also be a need for optimization of precursor flux within the target host. The metabolic engineering pathways for polyketide precursors must be integrated into hosts that express highly soluble and active engineered PKSs. By building on existing polyketide engineering and characterization strategies, harnessing current synthetic biology technologies and utilizing advances in metabolic/host engineering we anticipate future successes in engineering PKSs capable of producing designer polyketides for applications in medicine, fuels and industrial products.

2.7. Miscellaneous

2.7.1. Acknowledgements

This work was funded by the Joint BioEnergy Institute (JBEI), which is funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under Contract DE-AC02-05CH11231, by the National Science Foundation under awards MCB-1442724, NSF-GRFP DGE-1106400 and CBET-1437775, as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Offices, and by the DOE AgileBiofoundry (https://agilebiofoundry.org) supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy.

Bioenergy Technologies Office, through contract DE-AC02- 05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. Additional funding was provided by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE 1106400).

2.7.2. Competing interests

J.D.K. has financial interests in Amyris, Lygos, Constructive Biology, Demetrix, Napigen and Maple Bio.

Chapter 3. Structural Mechanism of Regioselectivity in an Unusual Bacterial Acyl-CoA Dehydrogenase

Including material from published work: **Blake-Hedges, J.M.**, Pereira, J.H., Cruz-Morales, P., Thompson, M.G., Barajas, J.F., Chen, J., Krishna, R., Chan, Leanne J.G., Nimlos, D., Alonso-Martinez, C., Baidoo, E.E.K., Chen, Y., Gin, J.W., Katz, L., Petzold, C.J., Adams, P.D., Keasling, J.D. Structural Mechanism of Regioselectivity in an Unusual Bacterial Acyl-CoA Dehydrogenase, *The Journal of the American Chemical Society*, **2019**, accepted https://doi.org/10.1021/jacs.9b09187

3.1. Abstract

Terminal alkenes are easily derivatized, making them desirable functional group targets for polyketide synthase (PKS) engineering. However, they are rarely encountered in natural PKS systems. One mechanism for terminal alkene formation in PKSs is through the activity of an acyl-CoA dehydrogenase (ACAD). Herein, we use biochemical and structural analysis to understand the mechanism of terminal alkene formation catalyzed by an γ,δ-ACAD from the biosynthesis of the polyketide natural product FK506, TcsD. While TcsD is homologous to canonical α,β-ACADs, it acts regioselectively at the γ,δ-position and only on α,β-unsaturated substrates. Furthermore, this regioselectivity is controlled by a combination of bulky residues in the active site and a lateral shift in the positioning of the FAD cofactor within the enzyme. Substrate modeling suggests that TcsD utilizes a novel set of hydrogen bond donors for substrate activation and positioning, preventing dehydrogenation at the α,β position of substrates. From the structural and biochemical characterization of TcsD, key residues that contribute to regioselectivity and are unique to the protein family were determined and used to identify other putative γ,δ-ACADs that belong to diverse natural product biosynthetic gene clusters. These predictions are supported by the demonstration that a phylogenetically distant homolog of TcsD also regioselectively oxidizes α,β-unsaturated substrates. This work exemplifies a powerful approach to understand unique enzymatic reactions and will facilitate future enzyme discovery, inform enzyme engineering, and aid natural product characterization efforts.

3.2. Introduction

Natural products often have complex chemical structures which can confer potent biological activity. Evolution selects for the diversification of these secondary

metabolites, making natural product biosynthetic pathways rich resources for the discovery of both lead compounds for drug discovery and also enzymes with unique functions.[207–209] Next-generation sequencing has led to the identification of numerous biosynthetic gene clusters (BGCs), but the pool of "orphan" BGCs (i.e. BGCs with no cognate natural product identified) remains largely untapped. Predicting natural product structures from primary DNA sequence is challenging, as the *in silico* functional annotation of enzymes within BGCs is limited. Amino acid sequence homology can suggest a general function, but without in depth structural and biochemical characterization of one or more members of an enzyme family, precise predictions of the final natural product structure are difficult. The identification of specificity-conferring motifs in polyketide synthase (PKS) acyltransferase (AT) and ketoreductase (KR) domains, for example, has allowed for more precise predictions of final polyketide natural product structure, including the identity of the alkyl substituents incorporated by (AT) and final stereochemical outcome (KR) of a given PKS module.[210] While much effort has been dedicated to elucidating signature motifs within PKS domains, many PKS-associated enzymes that generate less common functional groups (such as non-canonical starter and extender units) are not as well annotated or understood. Better characterization of the enzymes implicated in the biosynthesis of unique and reactive moieties would facilitate the engineering of novel polyketides with applications in medicine, in agriculture, or as commodity chemicals.[207,211,101]

One particular moiety of interest for PKS engineering is the alkene, as alkenes could easily be chemically derivatized to introduce a multitude of other desirable functionalities into a natural product.[119,212–223] Alkenes are often generated within a polyketide via the action of reductive domains, but they are sequestered within the polyketide backbone[8] and are therefore less sterically accessible for chemical modification than terminal alkenes.[224–232] Terminal alkenes, in addition to their innate reactivity, can also confer improved biological activity to polyketides that display biological activity, as is exemplified by the improved drug tolerability and efficacy of a synthetic epothilone analog sagopilone (anti-cancer activity).[233,234] There are few known examples of polyketides that contain terminal alkenes, including FK506, haliangicin, curacin A, and tautomycetin (**Figure 3-1A**).[235-236,237] In curacin A, a terminal alkene is generated through a unique off-loading mechanism involving a sulfotransferase (ST) and thioesterase domain.[118] The tautomycetin terminal alkene is known to be formed after the chain has been released from the PKS but the process has yet to be fully characterized.[238,239]

A different mechanism for terminal alkene formation occurs via the action of an acyl-CoA dehydrogenase (ACAD) that oxidizes the γ,δ-position of a fatty acyl-CoA or acyl carrier protein (ACP), as observed in the haliangicin and FK506 pathways.[240,241] The identification of other terminal alkene-forming ACADs is difficult, though, because of the enzymes' homology to canonical ACADs. Thus, identifying important sequence motifs is critical to more accurate annotation. ACADs are oxidoreductase flavoenzymes well known for catalyzing the first oxidative step of fatty acid β-oxidation: the dehydrogenation of saturated fatty acyl-Coenzyme A (CoA) thioesters to form the corresponding α,β-unsaturated product (**Figure 3-1B**).[242] The ACADs from the haliangicin and FK506 biosynthetic pathways, however, oxidize the γ,δ-position of a substrate. In order to facilitate better annotation of γ,δ-ACADs and, in turn, the identification of polyketide natural products that potentially contain terminal alkenes, a better functional characterization of γ,δ-ACADs is necessary. Herein, we report on our studies of the basis for the shift in regiochemistry of one of these unusual ACADs, TcsD, which forms the terminal alkene of the allylmalonyl-CoA extender unit in the biosynthesis of the polyketide FK506.[241]

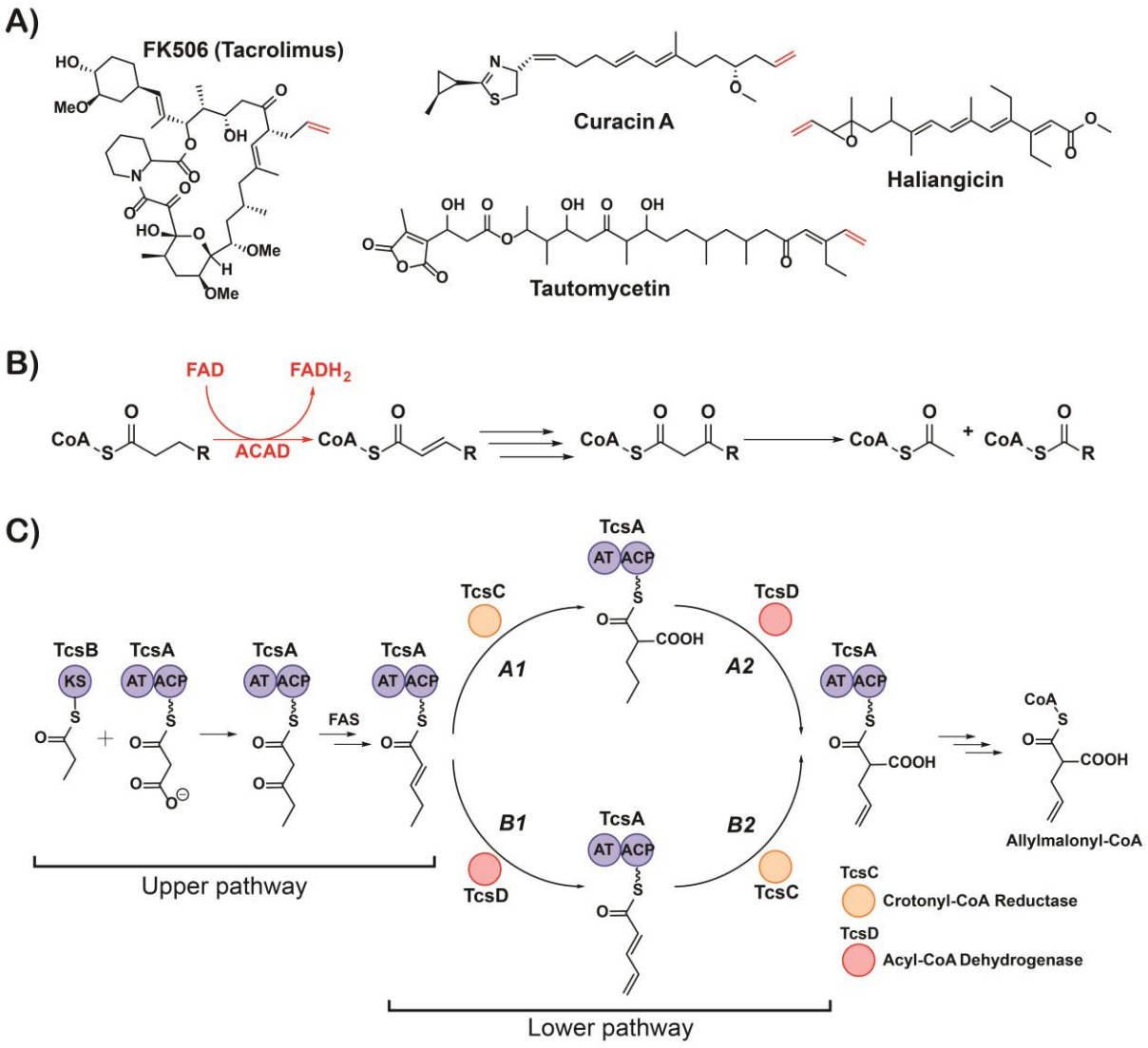


Figure 3-1 Terminal alkene-containing natural products and alkene formation by acyl-CoA dehydrogenases. **A)** Examples of terminal alkene-containing polyketide natural products. Terminal alkenes are highlighted in red. **B)** Process of fatty acid β-oxidation. The canonical activity of acyl-CoA dehydrogenases (ACADs) is the dehydrogenation of saturated fatty acyl-CoAs to form α,β-unsaturated products with concomitant reduction of FAD, highlighted in red. **C)** Proposed steps of allylmalonyl-CoA biosynthesis (adapted from reference 241). KS = ketosynthase domain, AT = acyltransferase domain, ACP = acyl carrier protein, FAS = fatty acid synthase.

The allylmalonyl-CoA biosynthetic pathway was initially proposed as a hybrid PKS-fatty acid synthase (FAS) pathway in which a free-standing ketosynthase, TcsB, first condenses a propionate group with a malonyl-CoA-derived extender unit selected by the AT domain of TcsA to form 3-oxo-pentanoyl-TcsA (**Figure 3-1C**, upper pathway).[241] After

reduction by the producing organism's FAS, it was proposed that the final two enzymes TcsC (a crotonyl-CoA reductase) and TcsD (an ACAD) work interchangeably to convert 2-pentenoyl-TcsA to allylmalonyl-TcsA, with TcsD forming the unique terminal alkene moiety (lower pathway).

The sequence homology that TcsD shares with other ACADs suggests that it utilizes the same chemical mechanism to form the γ,δ-alkene of the allyl functional group. ACADs employ a catalytic base, typically a glutamate residue, to deprotonate the acidic α-proton of a fatty acyl-CoA substrate.[242–245] The concomitant transfer of a hydride from the β-carbon of the substrate to FAD results in the formation of an α,β-unsaturated product and the reduced flavin, $FADH_2$.[246] The reaction is mediated by pK_a perturbations of both the catalytic glutamate residue and the substrate α-protons that occur within the active site of the enzyme. The pK_a of the glutamate is raised from ~6 to ~9 due to desolvation of the active site upon the binding of a hydrophobic substrate,[247–249] while the substrate protons are activated via hydrogen bonds of the substrate thioester carbonyl group with the amide backbone of the glutamate and the 2′-hydroxyl group of FAD.[250–253] While this mechanism of substrate activation is plausible for the γ-protons of a substrate such as 2-pentenoyl-TcsA (which contains an α,β-alkene that can propagate electronic effects from the thioester to the γ-carbon), the activation of propylmalonyl-TcsA and its conversion to allylmalonyl-TcsA (**Figure 3-1C**, pathway *A2*) is highly unlikely due to the aliphatic nature of the substrate.

Here we interrogate the activity of TcsD on both potential substrates and show that pathway *B1* (**Figure 3-1C**) is the only route of allylmalonyl-ACP formation. Additionally, we show that TcsD oxidizes only α,β-unsaturated substrates and is regioselective for the γ,δ position of these substrates. Further, we present a high resolution TcsD crystal structure and propose a structural mechanism by which it exhibits precise regiocontrol over this transformation. Combined structural and biochemical analyses of this enzyme revealed signature residues that contribute to its unique regioselectivity and facilitate the identification of homologs that display the same biochemical activity. A better understanding of this unique enzyme's activity will not only inform the characterization of other homologs and their associated BGCs, but the insights gained herein can also aid future enzyme engineering efforts.

3.3. Results

3.3.1. Biochemical Activity of TcsD

In order to understand the mechanisms underlying TcsD activity, we initially sought to determine the native substrate(s) of the enzyme by biochemically

interrogating both pathway *B*1 and *A*2 (**Figure 3-1C**). The substrates 2-pentenoyl-ACP (pathway *B1*) and propylmalonyl-ACP (pathway *A2*) were incubated with TcsD in the presence of the external electron acceptor ferrocenium hexafluorophosphate to facilitate enzyme turnover.[254] Assays were analyzed using targeted LC-MS/MS to monitor for the characteristic phosphopantetheine ejection transition (**Figure 3-2A**).[255] As expected, TcsD converted nearly all of the the 2-pentenoyl substrate to the corresponding 2,4-pentadienoyl-TcsA product, with no activity detected in boiled TcsD controls (**Figure 3-2B**). However, TcsD was unable to convert propylmalonyl-TcsA to allylmalonyl-TcsA under identical assay conditions (**Figure 4-6**). We therefore concluded that the biosynthesis of allylmalonyl-CoA can only proceed through route *B*, in which TcsD first oxidizes 2-pentenoyl-TcsA to 2,4-pentadienoyl-TcsA (*B*1) and subsequently TcsC performs a reductive carboxylation to form allylmalonyl-TcsA (*B*2).

Due to the noncanonical regiochemistry of the TcsD-mediated dehydrogenation of 2-pentenoyl-TcsA, we further investigated whether TcsD is regioselective for the γ,δ-position or if it simply oxidizes any appropriately activated substrate. It is known that some ACADs can abstract the acidic γ-proton of α,β-unsaturated substrates after dehydrogenation.[243,244,256] Accordingly, it is plausible that TcsD is a promiscuous enzyme that dehydrogenates any substrate present on TcsA-ACP and that any specificity it exhibits *in vivo* arises from the sequestration of substrates on a protein (TcsA) instead of Coenzyme A. We therefore probed TcsD activity on a panel of α,β-unsaturated and fully saturated substrates. The enzyme was active on another α,β-unsaturated substrate, 2-hexenoyl-TcsA, but inactive on the seven carbon 2-heptenoyl-TcsA (**Figure 3-2B**). On the substrates butyryl-TcsA and pentanoyl-TcsA, where possible α,β-unsaturation could be expected, no oxidation activity was observed which indicated that TcsD acts regioselectively at the γ,δ-position (**Figure 3-2B**, **4-7**, **4-8**).

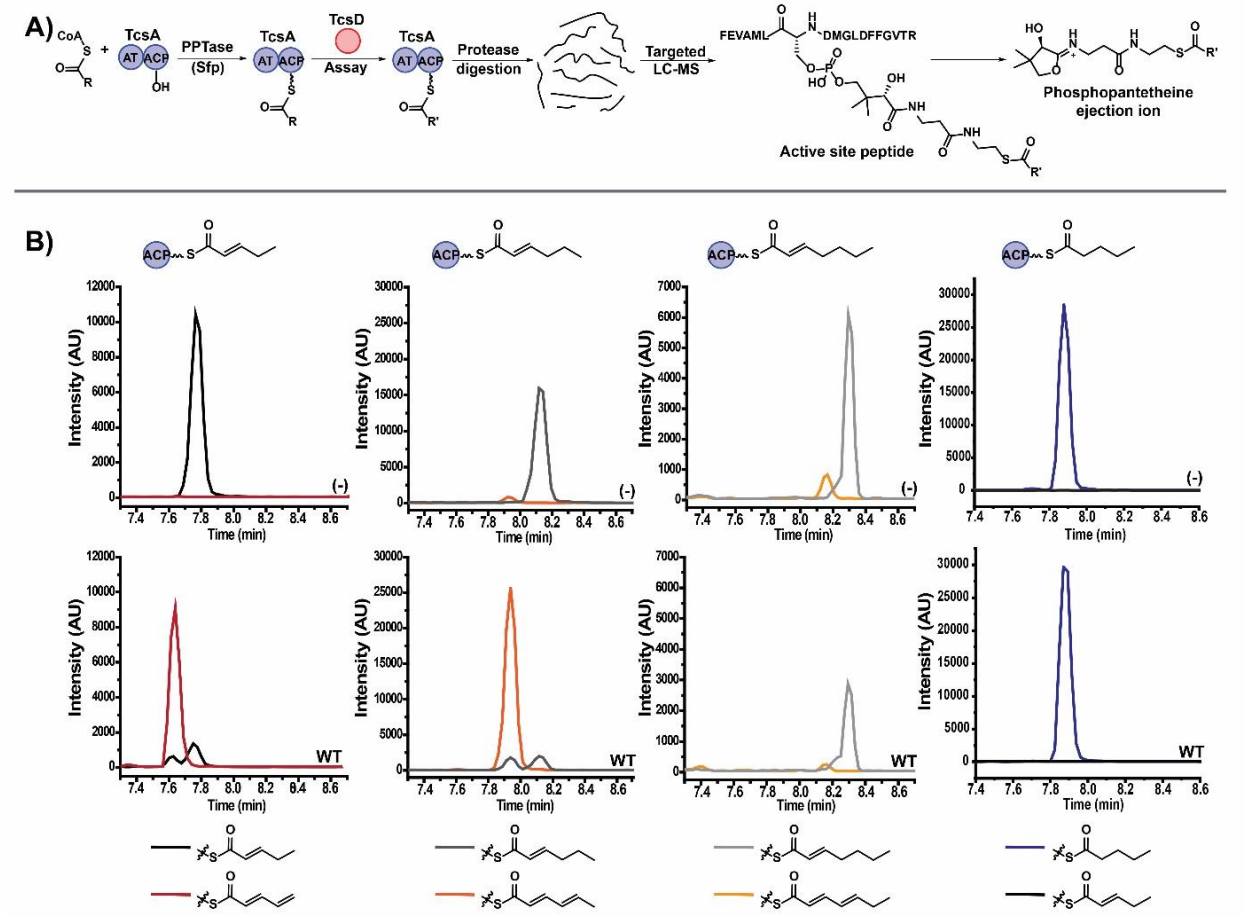


Figure 3-2 Biochemical activity of TcsD on ACP-bound substrates. A) Experimental design for phosphopantetheine ejection assays of TcsD activity on TcsA-bound substrates **B)** LC-MS/MS chromatograms of TcsD activity assays on various substrates. Substrates from left to right: 2-pentenoyl-TcsA, 2-hexenoyl-TcsA, 2-heptenoyl-TcsA, pentanoyl-TcsA. Top row: negative controls with boiled TcsD (-). Bottom row: assays with intact TcsD (WT).

3.3.2. Crystal Structure of TcsD

The strict substrate specificity of TcsD at the γ,δ-position suggested that there are structural elements within the enzyme's active site that control regioselectivity. To test this hypothesis, we obtained a crystal structure of TcsD, which was solved to 1.75Å resolution. The TcsD structure displays many similarities to those of other ACADs, such as the conserved ACAD fold consisting of a tetrameric quaternary structure, which is further organized into two sets of homodimers (**Figure 4-9**). Each subunit contains a single FAD cofactor and is composed of 3 subdomains consisting of a set of N- and C-terminal alpha helix domains that surround a middle beta sheet domain. The FAD cofactor adopts an extended conformation and is situated in a pocket formed by the C-

terminal alpha helix domain, the middle beta sheet domain, and the C-terminal domain of the adjacent subunit (**Figure 4-9**).

The active site of TcsD also shares a similar overall organization with α,β-ACADs. Like the structures of the human, rat, pig, and bacterial ACADs, the catalytic base of TcsD, Glu364, sits poised at the top of the active site pocket immediately above the fatty acyl binding chamber.[242,245,257–259] The back side of the pocket that bounds the fatty acyl binding region is shaped by the residue immediately upstream of the catalytic glutamate, Ile363, and three residues from helix 5: Phe79, Leu83, and Leu86. The FAD cofactor sits at the bottom of the active site, positioned below Glu364 (**Figure 3-3A**).

The isoalloxazine ring of FAD is anchored via conserved hydrogen bonds with residues Thr116, Gly118, Ser119, and Thr151, while the adenosine pyrophosphate portion of the molecule extends into a cavity formed between the loop that follows β-sheet 1 and helices 10 and 11 of the adjacent subunit; it is positioned through polar interactions with Ser125 and Glu337 of the same subunit and Met338, Gly340, and Gly341 of the adjacent monomer (**Figure 4-10**). Similar to other ACAD structures, an aromatic residue, Phe149, is positioned on the si face of the isoalloxazine ring. The FAD is positioned adjacent to the substrate binding cavity which is bound on the opposite side by several conserved residues from helix 7 and the loop following sheet 1, including Gly118, Ser119, Glu120, Ser125, Leu127, Leu224, and Leu228 which form the phosphopantetheine binding region.

Although similar in overall structure, TcsD possesses many features that distinguish it from canonical ACADs. The residues surrounding the TcsD active site entrance have diverged from the corresponding residues that are conserved in α,β-ACADs and contribute to the positioning of the adenosine portion of the coenzyme A substrate. TcsD does not possess the conserved Asp and Thr residues that form hydrogen bonds with the adenine base of Coenzyme A (**Figure 4-11**), nor the small helix between β-sheets 5 and 6 that often hydrogen bonds with the coenzyme A phosphate groups (**Figure 4-12**). The loop between β-sheets 5 and 6 in TcsD has instead been shortened. Additionally, a short helix, helix 9, is present in the TcsD structure in a position that would collide with the

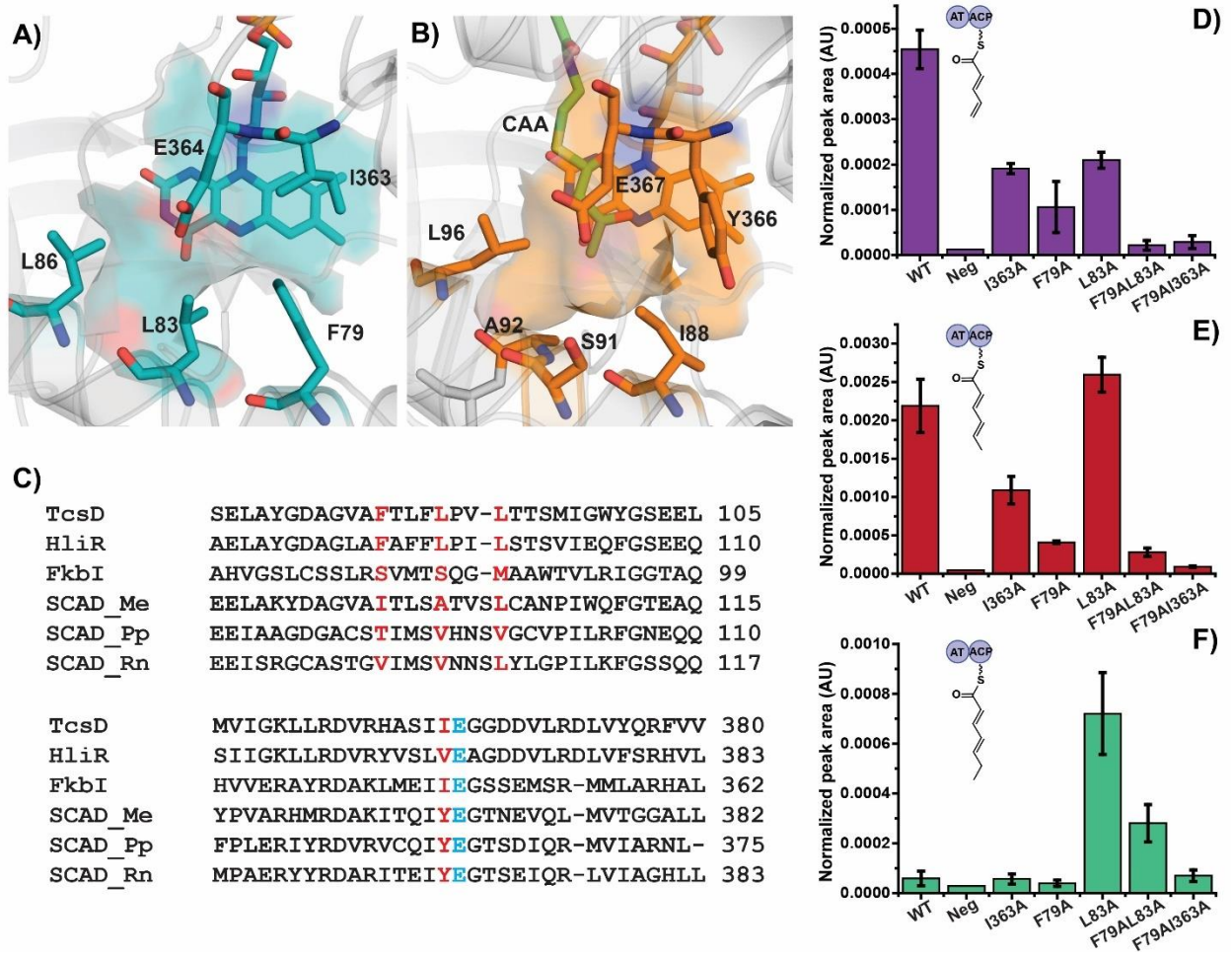


Figure 3-3 Unique active site features of TcsD and activity of active site mutants. A) Shape of the fatty acyl binding region of TcsD. **B)** Shape of the fatty acyl binding region of the *Megasphaera elsdenii* butyryl-CoA dehydrogenase (pdb entry 1buc), which was co-crystallized with acetoacetyl-CoA (CAA) **C)** Sequence alignment of TcsD with HliR,[237] FkbI,[260] *M. elsdenii* butyryl-CoA dehydrogenase (SCAD_Me, 1buc),[259] *Pseudomonas putida* KT2440 short chain acyl-CoA dehydrogenase (SCAD_Pp),[261] and rat short chain acyl-CoA dehydrogenase (SCAD_Rn, 1jqi).[258] The residues in red correspond to residues that form the acyl-binding region of the enzyme active site (Phe79, Leu83, Leu86, and Ile363 in TcsD). The active site glutamate is highlighted in blue. **D)**, **E)**, **F)** TcsD mutant activity on 2-pentenoyl-TcsA* (D), 2-hexenoyl-TcsA* (E), and 2-heptenoyl-TcsA* (F) calculated as the LC-MS/MS peak area of the product normalized to a control peptide from TcsA (Normalized peak area, see Methods in SI). Data displayed are the average of three replicates. AU = arbitrary units *Note: TcsD activity on 5 carbon substrates cannot be compared to TcsD activity on 6 or 7 carbon substrates.

adenine base of Coenzyme A as it is positioned in α,β-ACAD structures. While helix 9 in the TcsD structure seems to interfere with Coenzyme A binding, it is also present in the

TcsD homolog from the haliangicin pathway, HliR, suggesting that this helix might contribute to the regiochemical shift observed in both enzymes (**Figure 4-13**).

TcsD has unique structural features that form the fatty acyl binding region of the active site pocket. In particular, Ile363 and the residues of helix 5 that line the back side of the active site pocket protrude further into the active site than other ACADs, forming a much shallower binding pocket (**Figure 3-3A** & **3-3B**). More specifically, Phe79 and Leu83 protrude directly into the active site pointing toward the fatty acyl binding position. In addition, these two residues are bulkier than the residues forming the active site of canonical ACAD homologs, contributing to the reduced active site pocket depth seen in the TcsD structure. The difference in active site shape and depth is further illustrated by comparing TcsD with the structures and sequences of α,β-ACADs containing co-crystallized substrates. Alignment and superposition of the *Megasphaera elsdenii* butyryl-CoA dehydrogenase (BCAD) in complex with acetoacetyl-CoA[259] with the TcsD active site shows that Phe79, Leu83, Leu86, Ile363 or a combination of these residues would sterically clash with substrates longer than acetoacetyl-CoA if they were to bind within the TcsD active site in a similar manner. The *M. elsdenii* BCAD and other short chain α,β-ACADs contain less bulky residues at the positions corresponding to Phe79, Leu83, and Ile363, but Leu86 is generally conserved in TcsD, HliR, and α,β-ACADs (**Figure 3-3B** & **4-13**).[71,237,258,260,261] In addition, the bulky residues Phe79 and Leu83 distinguish TcsD from FkbI, the hydroxymalonate semialdehyde dehydrogenase from the methoxymalonyl-ACP biosynthetic pathway of FK506 which has serines at both of these positions.[260]

Given the unique structural characteristics of TcsD, we hypothesized that the positioning of the amino acids Phe79, Leu83, and Ile363 could control the regioselectivity of TcsD by preventing either proton abstraction by Glu364 or hydride transfer to FAD through steric repulsion. More specifically, these large residues would prohibit a five-carbon substrate from entering the active site far enough so that Glu364 can access the protons bound to the substrate α-carbon or for the β-carbon to transfer a hydride to N5 of FAD. Rather, the substrate would be pushed towards the active site entrance, positioning Glu364 above the γ-carbon and the δ-carbon within an appropriate distance of N5 of FAD. As the substrates pentanoyl-ACP and 2-pentenoyl-ACP are similarly chemically activated (the pK_as of the α- and γ-protons are similar), the regioselectivity of TcsD would be controlled purely by steric interactions, not an electronic preference for one substrate.

3.3.3. Biochemical activity of TcsD mutants

In order to verify whether the regioselectivity of TcsD is sterically controlled, a mutant of TcsD with a larger active site pocket that can accommodate a longer substrate is required. Specifically, TcsD mutants that act upon a substrate with two additional carbons should also be able to bind pentanoyl-ACP in a manner that properly positions the α- and β-carbons within an accessible distance of Glu364 and FAD, respectively, as the α-carbon is two carbons removed from the γ-carbon. The residues that surround the active site of TcsD were therefore selectively mutated to alanines in order to accommodate longer substrates. Individual alanine mutants of Phe79, Leu83, and Ile363 in addition to double mutants of neighboring residues (e.g. F79A/L83A and F79A/I363A) were generated, and the dehydrogenation activity of mutants was then probed on a panel of α,β-unsaturated substrates. The active site mutants were generally less active on 2-pentenoyl-ACP than the wild type enzyme (**Figure 3-3D**). However, the I363A and L83A mutants displayed nearly equal or equivalent activity to the wild type on 2-hexenoyl-ACP (**Figure 3-3E**). Proteins containing the L83A mutation were also active on the 2-heptenoyl-ACP substrate (**Figure 3-3F**). This indicates that Leu83 controls the chain length of the substrate.

After demonstrating that the active site of the L83A mutant had been enlarged to accommodate two additional carbons, we next probed its activity on 4- and 5-carbon fully saturated substrates to test if the enzyme was now capable of dehydrogenating the α,β-position in the absence of the wild type steric interactions. We found that the mutant enzyme was still inactive on butyryl- and pentanoyl-ACP (data not shown), suggesting that the observed regioselectivity is not controlled exclusively by steric interactions with the fatty acyl tail of a substrate.

3.3.4. Substrate modeling into TcsD active site

Given the retention of selectivity of the TcsD L83A mutant, we hypothesized that the regioselectivity of dehydrogenation could instead be controlled by intermolecular forces affecting the positioning of the thioester end of the substrate, which is determined by two hydrogen bonds (H-bonds) within the active site. We hypothesized that the canonical ACAD H-bonds should be conserved in TcsD as they are crucial to the chemical mechanism of the enzyme. To test this, we attempted to co-crystallize TcsD with 2-pentenoyl-CoA and 2-hexenoyl-CoA in order to show the substrate positioning within the active site. However, we were unable to observe density corresponding to either substrate in the TcsD active site pocket, possibly because an unknown *E. coli* metabolite consistently co-crystallized in the active site and occupied the space where a substrate should bind (**Figure 4-3**).

Instead, we computationally modeled the native TcsD substrate into the active site to understand which structural features might affect substrate binding. The positioning of the thioesters in the structures of several ACADs (which were co-crystallized with substrate analogs) was used to guide the placement of a substrate into the TcsD active site. More specifically, the structures of several homologs and TcsD were superimposed by aligning the peptide backbones of the three-residue loop that contains the catalytic glutamate (e.g. Ile363, E364, and G365). This three-residue loop was chosen as an anchor point because it is highly conserved and contains one of the H-bond donors that contributes to thioester positioning, so the relative positioning of structural elements within the ACAD active sites can be used to infer relative substrate positioning. Not only was this type of alignment useful for appropriately positioning modeled substrates, but it also allowed for a more precise determination of slight structural differences that were not as obvious when analyzing global structural alignments. In particular, it highlighted stark differences in the positioning of the FAD cofactor in TcsD relative to several homologs (**Figure 3-4A** & **3-4B**).

In relation to the active site glutamate (Glu364), the FAD molecule bound within the TcsD active site is shifted laterally, moving it in the direction of the pantetheine-binding region of the enzyme (**Figure 3-4A** & **3-4B**). This movement of the ribityl side chain of FAD is accompanied by a shift in α-helices 10 and 11 and the loop that follows β-sheet 1, which interact with the FAD through various H-bonds. The repositioning of α-helices 10 and 11 results in a change in the conformation of the FAD ribose ring due to H-bonding interactions of the 2′ and 3′-OH groups with the amide backbone of Met338 and of Glu337 and Gly339 of the adjacent subunit, respectively. The conformational change pushes the FAD phosphate groups towards the β-sheet domain of the enzyme. As a result of these structural changes, the positioning of the 2′-hydroxyl group of the FAD ribityl side chain is shifted closer to the nitrogen of the Glu364 amide bond (**Figure 3-4A**). We hypothesized that this would result in a concomitant shift in the positioning of the substrate thioester moiety, making TcsD act at the γ,δ-position instead at the canonical α,β-position of substrates.

In order to better understand how the change in FAD positioning affects substrate binding, we modeled a substrate analog consisting of a 2-pentenoyl thioester into the active site. Initially, we positioned the thioester carbonyl oxygen within hydrogen bonding distance of both the 2′-OH of the FAD and the nitrogen of the amide bond of Glu364 (**Figure 3-4C**). Anchoring the carbonyl group in this location would force

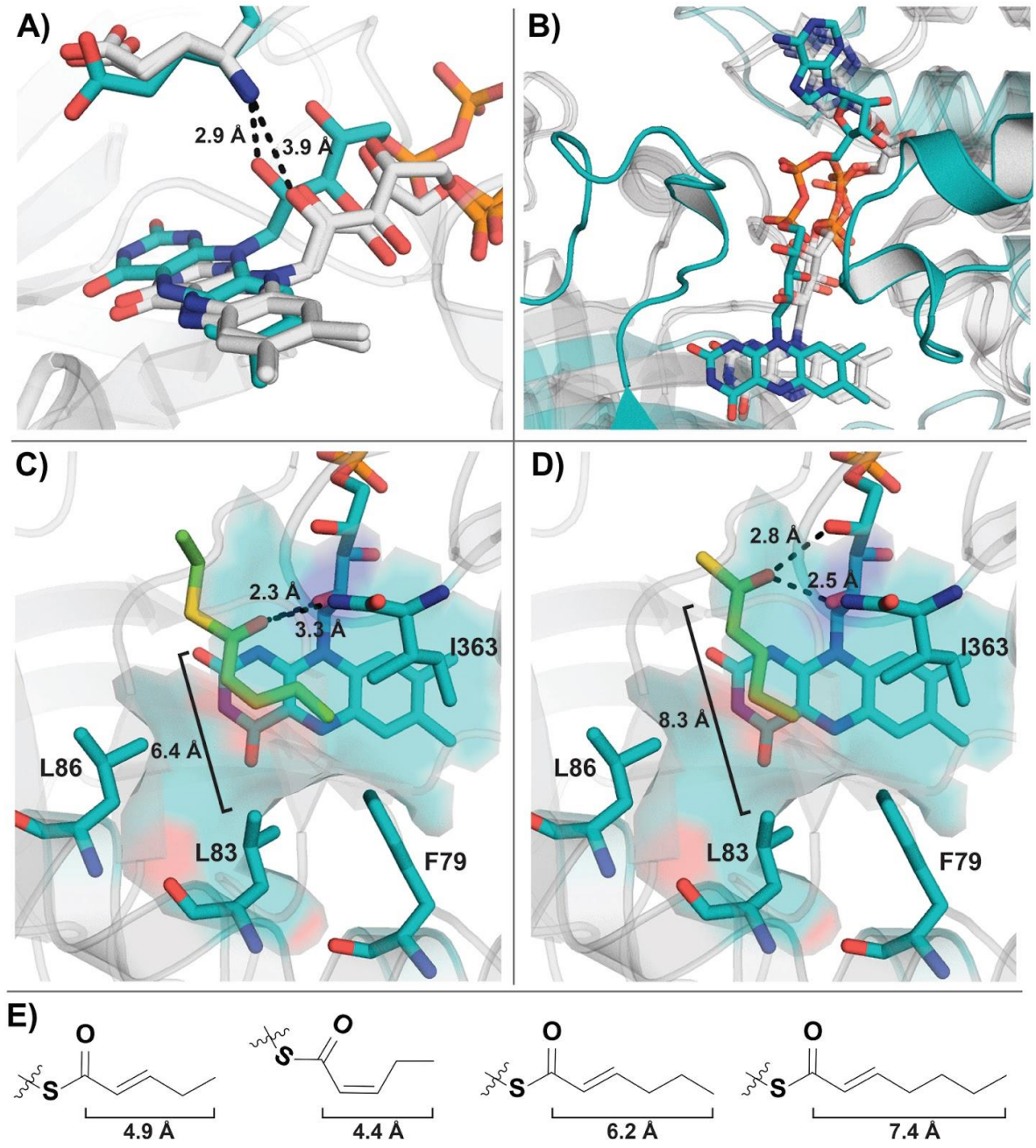


Figure 3-4 FAD shift and substrate modeling in TcsD active site. Shift in positioning of the 2′-OH group of FAD in TcsD (teal) relative to the active site glutamate and compared to the pig (*Sus scrofa*) medium chain ACAD in the apo (pdb code 3mdd)[43] and substrate-bound (pdb code 1udy)[55] forms (both white). Dotted lines represent distances between 2′-OH groups of FAD and amide nitrogens of active site glutamates. **B)** Lateral shift of FAD in TcsD (teal) with respect to the active site glutamate and compared to the pig (*Sus scrofa*) medium chain ACAD in the apo (pdb code 3mdd) and substrate-bound (pdb code 1udy) forms (both white) **C)** TcsD active site* containing a modeled 2-pentenoyl thioester group in the *cis* conformation. Dotted lines = distances between the substrate thioester carbonyl group and the 2′-OH of FAD and amide

nitrogen of Glu364 **D)** TcsD active site* containing a modeled 2-pentenoyl thioester group in the *trans* conformation. Dotted lines = distances between the substrate thioester carbonyl group and the 2′-OH and 4′-OH of FAD **E)** Distance from carbonyl carbon to the tail carbon of various α,β-unsaturated fatty acyl substrates. *Note: Glu364 was omitted from C) and D) in order to visualize the hydrogen bonding interactions of the substrate with FAD.

the substrate alkene into a *cis* conformation, as this is the only conformation that places the δ-carbon in close enough proximity to N5 of FAD. However, while this conformation enforces the proper positioning of the substrate with respect to FAD, it places the δ-carbon too close to residues Phe79 and Ile363 which would result in unfavorable steric repulsion between the substrate and the amino acid side chains lining the active site. Moreover, with this substrate conformation there is no additional space within the active site to accommodate a sixth carbon on the substrate which conflicts with our biochemical data.

After eliminating the possibility of substrate binding in the *cis* conformation, a *trans* substrate was next modeled into the TcsD active site. When the substrate alkene bond adopts a *trans* conformation, its positioning is more consistent with the biochemical activity of the enzyme as it places the fatty acyl tail turned toward Leu83 and the δ-carbon in proximity to the N5 of FAD (**Figure 3-4D**). However, with the substrate in this extended position, H-bonding of the substrate with the canonical ACAD H-bond donors cannot occur without a steric clash between Leu83 and the fatty acyl tail of the substrate. We therefore modeled the carbonyl carbon of the substrate so that it H-bonds with the 2′ and 4′-OH groups of FAD, not with the canonical amide nitrogen of Glu364. Given the size dimensions of the active site and the biochemical data presented herein, this appears to be the most probable positioning of the substrate within the active site, suggesting that TcsD utilizes a novel H-bond donor pair to achieve its unique regioselectivity.

3.3.5. Genome mining reveals previously unidentified γ,δ-ACADs

After identifying structural features that contribute to the unique regioselectivity of TcsD, we applied this mechanistic knowledge to find unidentified or misannotated γ,δ-ACADs among sequenced bacteria. In particular, we used the presence of bulky residues at positions corresponding to Phe79, Leu83, and Ile363 and helix 9 as requirements for the identification of γ,δ-ACADs. With these constraints, approximately 100 likely γ,δ-ACADs were identified using a combination of Hidden Markov Model (HMM), local sequence alignment (protein BLAST), and CORASON-BGC searches (**Table 3-9** & **4-1**).[262–264]

Notably, both TcsD and HliR, the only γ,δ-ACADs associated with known natural products, were re-identified in our search. Of the newly-identified enzymes, one feature in particular was strongly conserved: the presence of helix 9, which we had identified as a unique feature in the TcsD structure (Figure 5A and B). The exact purpose of this loop remains unclear and will be the subject of further investigation. Additionally, we found that the presence of bulky residues at positions in the substrate acyl binding region were highly conserved across the family, with 98% of homologs displaying a Phe residue (2% Leu) and 96% displaying a Leu or Ile residue (4% Val) at the positions corresponding to Phe79 and Leu83 in TcsD (Figure 5C and S6). The residues immediately preceding the catalytic glutamate (i.e. residue Ile363 in TcsD) were all bulky aliphatic residues as well (87% Val, 11% Ile, 2% Phe). However, bulky residues at this position are not unique to the γ,δ-ACAD family as they are also observed in hydroxymalonate semialdehyde dehydrogenases such as FkbI.

Most of the TcsD homologs were associated with identifiable BGCs, although some were not located near any obvious secondary metabolite genes. The majority of the BGC-associated genes were located within type I PKS (T1PKS) clusters, but several were also identified within type II PKS (T2PKS) and NRPS clusters (**Figure 3-5D**, **4-15** & **4-16**). None of the newly-identified BGCs containing the putative γ,δ-ACADs have been experimentally characterized, but many show significant homology to known clusters (**Table 4-1**). Of the clusters that include putative γ,δ-ACADs, several contained proteins with homology to BGCs with known products, including the arsenopolyketides (T1PKS), oligomycin (T1PKS), E-837 (T1PKS), chlorothricin (T1PKS), butyrolactols (T1PKS), polyoxypeptin (T1PKS-NRPS), hedamycin (T2PKS), and erythrochelin (NRPS). Many of the homologous BGCs identified by antiSMASH have fatty acyl tails that could be potential substrates for γ,δ-ACADs in the uncharacterized clusters we identified.

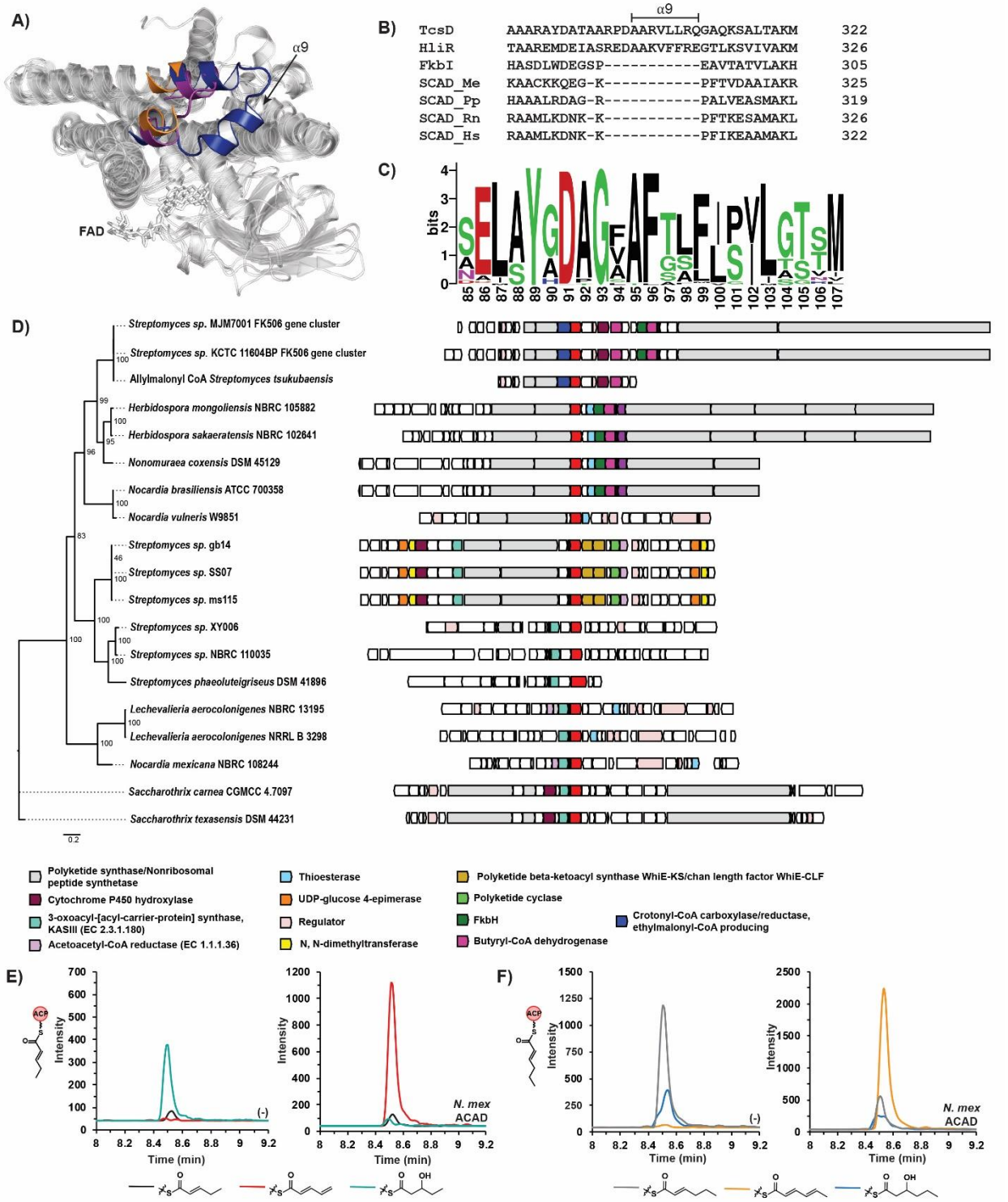


Figure 3-5 Genome mining and characterization of γ,δ-ACADs. A) Global overlay of TcsD with *M. elsdenii* butyryl-CoA dehydrogenase (BCAD_Me, 1buc)[259] and *Streptomyces hygroscopicus* hydroxymalonate semialdehyde dehydrogenase, FkbI (1r2j)[260] highlighting the insertion of helix 9 into γ,δ-ACAD structures. The colored regions are helix 9 of TcsD (blue, labeled *α*9) and the corresponding regions in BCAD_Me

(purple) and FkbI (orange). FAD is depicted as sticks. **B)** Sequence alignment of TcsD with HliR,[237] FkbI, *M. elsdenii* butyryl-CoA dehydrogenase (SCAD_Me, 1buc), *Pseudomonas putida* KT2440 short chain acyl-CoA dehydrogenase (SCAD_Pp),[261] and rat short chain acyl-CoA dehydrogenase (SCAD_Rn, 1jqi).[258] The sequence alignment highlights the presence of helix 9 in γ,δ-ACADs which is not encountered in other ACAD families. **C)** Amino acid position weight matrix showing the relative abundance of amino acids at each position within γ,δ-ACADs. Positions 96 and 100 correspond to the positions of Phe79 and Leu83 of TcsD, respectively. Logo was generated from a sequence alignment of all γ,δ-ACADs identified in this work. **D)** Genomic contexts and synteny of a representative set of γ,δ-ACADs identified in this work.TcsD homologs (putative γ,δ-ACADs) are shown in red and aligned. Other notable genes are annotated by color. **E)** Targeted LC-MS/MS chromatograms showing the activity of the *Nocardia mexicana* NBRC 108244 putative γ,δ-ACAD on 2-pentenoyl-ACP **F)** Targeted LC-MS/MS chromatograms showing the activity of the *Nocardia mexicana* NBRC 108244 putative γ,δ-ACAD on 2-hexenoyl-ACP. (-) is used to mark negative controls with boiled *N. mexicana* ACAD, while assays are marked as *N. mex* ACAD. Chromatograms representing the 2,3-enoyl-, 2,4-dienoyl-, and 3-hydroxy transitions are color-coded below each set of chromatograms. The 3-hydroxyacyl groups are normal degradation products of 2,3-enoyl-thioesters in aqueous environments where hydration of the alkene can occur.

Within the identified BGCs, analysis of the genomic contexts of the putative γ,δ-ACADs showed several patterns of syntenic genes (**Figure 3-5D**, **4-15**, & **4-16**). The syntenic genomic regions can be grouped based on the type of gene cluster (e.g., probable butyrolactol or arsenopolyketide BGC) within which the putative γ,δ-ACADs occur. The groups of genes that are found near γ,δ-ACADs can also be used to postulate the enzymes' native substrates and even certain aspects of the molecular structure of the final natural product. The γ,δ-ACADs found in the erythrochelin-like gene clusters in *Saccharothrix sp.*, for example, are consistently clustered with a free-standing ketosynthase (KS) and acyl carrier protein (ACP) pair (**Figure 3-5D**). It is plausible that the KS domain forms a 5 carbon ACP-bound substrate upon which the γ,δ-ACAD can act after it is reduced to a 2-pentenoyl form. The genomic context of this and many of the other putative γ,δ-ACADs can be used to infer function and will inform future biosynthetic studies on the pathways and enzymes.

The presence of conserved residues and helix 9 in the putative γ,δ-ACADs is highly suggestive that they would exhibit the same activity as TcsD. To interrogate this hypothesis, we biochemically characterized a TcsD homolog. We chose the putative γ,δ-ACAD from *Nocardia mexicana* NBRC 108244 because it is located on a distant branch of the γ,δ-ACAD phylogenetic tree (**Figure 3-5D**, **4-15**, & **4-16**) and has low sequence identity relative to TcsD (50.7%). Though sharing only 51% sequence identity, the *N. mexicana* homolog (Nmex-ACAD) possesses the conserved Phe, Leu, and Ile residues that line the TcsD active site (**Figure 4-17**). It is located within a gene cluster that is predicted to encode a ladderane/butyrolactone-like natural product BGC (**Figure 4-18**).

There are several related biosynthetic genes located immediately next to the Nmex-ACAD, including an acyl carrier protein, a ketosynthase, and a putative 3-oxoacyl-ACP reductase. Based on its genomic context we hypothesized that, like TcsD, the Nmex-ACAD may act on ACP-bound substrates. We therefore expressed and purified both the Nmex-ACAD and its neighboring acyl carrier protein (Nmex-ACP) and assayed Nmex-ACAD activity on the same panel of substrates. The Nmex-ACAD showed the same activity profile as TcsD, converting 2-pentenoyl-Nmex-ACP (Figure 5E) and 2-hexenoyl-Nmex-ACP (**Figure 3-5F**) to the corresponding dienoyl-ACP products, but it showed no activity on butyryl-, pentanoyl-, or 2-heptenoyl-ACP (**Figure 4-19**). While this data does not confirm that all the putative enzymes we have identified are γ,δ-ACADs, it strongly supports the prediction that these enzymes have the same activity as they are more closely phylogenetically related to TcsD than the Nmex-ACAD and share the conserved bulky residues that form the enzyme fatty acyl binding pocket.

3.4. Conclusion

Terminal alkenes in polyketide natural products can be formed in several ways, including through the action of an acyl-CoA dehydrogenase. In this work we have described the biochemical and structural characterization of TcsD, the terminal alkene-forming γ,δ-ACAD from the biosynthesis of the allylmalonyl-CoA extender unit implicated in the biosynthesis of the polyketide FK506. We showed that TcsD acts on 2-pentenoyl-TcsA but not on propylmalonyl-TcsA, suggesting that the bottom half of the allylmalonyl-CoA pathway proceeds only through pathways *B1* and *B2*. TcsD only acts on 5- and 6-carbon α,β-unsaturated substrates and is regioselective for the γ,δ-position.

A crystal structure of TcsD revealed the unique features of the active site of the enzyme. Residues Phe79, Leu83, and Ile363 form a bulky wall in the substrate binding region of the enzyme, preventing the entrance of long fatty acyl substrates. Leu83 controls the chain length of the substrate. A TcsD L83A mutant acts on 2-heptenoyl-TcsA, but even with a larger active site pocket the mutant remains regioselective for the γ,δ-position of substrates. Closer analysis of the protein structure revealed that the enzyme regioselectivity is likely due to a shift in the positioning of the FAD cofactor. We show through substrate modeling that, because of the FAD shift and the dimensions of the TcsD active site, TcsD most likely employs a novel hydrogen bond donor pair (the 2′-OH and 4′-OH groups of FAD) to position and activate substrates. While Leu83 does not exclusively control regioselectivity, it contributes to regioselectivity by reducing the size of the active site.

The structural and biochemical conclusions from biochemically and structurally characterizing TcsD allowed us to determine key residues that define γ,δ-ACADs. Through HMM and local alignment searches, approximately 100 putative γ,δ-ACADs were identified in sequenced bacterial genomes. Nearly all of the homologs contained a Phe-Leu/Ile pair at the positions corresponding to Phe79-Leu83 in TcsD, respectively. The identification of other homologs also highlighted the conservation of helix 9, which appears to be a feature that is unique to the γ,δ-ACAD family. Most of the homologous enzymes were encountered in identifiable secondary metabolite BGCs, but some were notably located near no other canonical specialized metabolic enzymes. The synteny of genes located near the γ,δ-ACADs and the type of BGC to which they belong correlates strongly with their phylogenetic clustering and can be used to group the enzymes into several subfamilies. Finally, we showed that one of the TcsD homolog from *Nocardia mexicana*, which is phylogenetically one of the most distant enzymes from TcsD, also performs a regioselective dehydrogenation of the γ,δ-carbon of 2-pentenoyl- or 2-hexenoyl-ACP substrates, suggesting that this activity is conserved across the entire family.

This work exhibits how selective pressure causes enzymes to diverge not only at the amino acid sequence level, but also how it can result in significant shifts in protein structure to generate enzymes with divergent functions. Furthermore, it exemplifies how an understanding of the mechanisms employed by unique enzymes can be used as a means to refine the definitions of enzyme families and identify uncharacterized homologs. It will inform future efforts to characterize the identified homologs and the BGCs they residue within and can be used as a guide for the future discovery of natural products that contain terminal alkene handles.

3.5. Materials and methods

3.5.1. Materials, reagents, and strains

All strains used and generated in this work are listed in **Table 3-2**. Routine *E. coli* cultures were grown at 37°C in Luria-Bertani (LB) Miller medium (BD Biosciences, USA) and were supplemented with kanamycin (50 mg/L, Sigma Aldrich, USA), or carbenicillin (100mg/L, Sigma Aldrich, USA). *Pseudomonas putida* KT2440 was grown at 30°C in Luria-Bertani (LB) Miller medium (BD Biosciences, USA), and *Streptomyces tsukubaensis* NRRL 18488 was grown at 30°C in BD Bacto tryptic soy broth (TSB) medium (Fisher Scientific). All chemicals and reagents were purchased from Sigma Aldrich unless otherwise noted.

3.5.2. DNA Manipulation

All plasmids used and generated in this work are listed in **Table 3-1**. The genomic DNA samples from *Pseudomonas putida* KT2440 were purified using a DNeasy Blood and Tissue Kit (Qiagen, USA). Plasmids and PCRs were routinely isolated using the Qiaprep Spin Miniprep kit (Qiagen, USA) and DNA Clean & Concentration Kit (Zymo Research, USA). All primers were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Constructs were generated using restriction enzyme cloning (TcsD, TcsA wildtype coding sequences), Gibson assembly (TcsD mutants, PP2216), or Golden Gate assembly (*N. mex* ACAD and ACP) as previously described.[265,266]

3.5.3. Construct cloning: TcsD and TcsA

Genomic DNA of *Streptomyces tsukubaensis* NRRL 18488 was prepared by resuspending cells grown in TSB medium in LC-grade DMSO. The coding sequences of TcsD and TcsA were amplified from this genomic DNA sample using Q5 Hotstart high-fidelity polymerase using primers which incorporated flanking NdeI and XhoI restriction sites. The genes were then ligated into a pET28a backbone which had been linearized with the FastDigest enzymes NdeI and XhoI (Thermo Scientific, USA), appending a C-terminal 6x histidine tag onto each. Targeted mutations were generated in TcsA and TcsD using Golden Gate cloning. Mutations were embedded in the overlap regions of the Golden Gate primers listed in **Table 3-3.**

3.5.4. Construct cloning: PP2216

PP2216 was amplified from *Pseudomonas putida* KT2440 genomic DNA and cloned into a pBbE7a backbone with a C-terminal 6x histidine tag using Gibson assembly.

3.5.5. Construct cloning: *Nocardia mexicana* ACAD and ACP

PThe coding sequences for the *Nocardia mexicana* NBRC 108244 putative γ,δ-acyl-CoA dehydrogenase (GenBank accession WP_068019852.1) and the neighboring acyl carrier protein (GenBank accession WP_068019854.1) were reverse translated and codon-optimized for *E. coli* using BOOST.[267] Genes were designed with flanking Golden Gate sites and purchased from Genscript (Piscataway, NJ, USA). The genes were then inserted into a pBbE7a backbone with a C-terminal 6x histidine tag using Golden Gate assembly. The sequences of the synthetic gene fragments are listed in **Table 3-4**.

Table 3-1 Plasmids used in this study

ICE Entry Number	Plasmid name	Description	Source
JPUB_013738	pET28a-TcsA	TcsA in pET28a	This study
JPUB_013740	pET28a-TcsAS98A	TcsA S98A in pET28a	This study
JPUB_013728	pET28a-TcsD	TcsD in pET28a	This study
JPUB_013730	pET28a-TcsDL83A	TcsD L83A in pET28a	This study
JPUB_013732	pET28a-TcsDF79A	TcsD F79A in pET28a	This study
JPUB_013734	pET28a-TcsDI363A	TcsD I363A in pET28a	This study
JPUB_013736	pET28a-TcsDF79AL83A	TcsD F79AL83A in pET28a	This study
JPUB_013744	pET28a-TcsDF79AI363A	TcsD F79AI363A in pET28a	This study
JPUB_013742	pBbE7a-PP2216-6xHis	PP2216 in pBbE7a with a C-terminal 6xHis tag	This study
JPUB_013842	pBbE7a-NmexACP-6xHis	Acyl carrier protein from *Nocardia mexicana* NBRC 108244 (*E. coli* codon-optimized) in pBbE7a with a C-terminal 6xHis tag	This study
JPUB_013844	pBbE7a-NmexACAD-6xHis	Acyl-CoA dehydrogenase from *Nocardia mexicana* NBRC 108244 (*E. coli* codon-optimized) in pBbE7a with a C-terminal 6xHis tag	This study

N/A	pET-Sfp	Sfp in a pET vector	Reference 169
N/A	pLK54	MatB mutant	Reference 150
N/A	pBbE7a-RFP	RFP behind a T7 promoter in a BglBrick plasmid	Reference 289

Table 3-2 Strains used in this study

ICE Entry Number	**Strain name**	**Description/Genotype**	**Source**
N/A	*E. coli* DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^-	QB3 Macrolab (http://qb3.berkeley.edu/macrolab/)
N/A	*E. coli* NEB Turbo	F' *proA⁺B⁺ lacI^q ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*Tet^S *endA1 thi-1 Δ(hsdS-mcrB)5*	New England Biolabs
N/A	*E. coli* BL21 (DE3)	F⁻ *ompT gal dcm lon* $hsdS_B(r_B^- m_B^-)$ λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB⁺*]$_{K-12}$(λ^S)	New England Biolabs
N/A	*Streptomyces tsukubensis*	Wild type strain	US Department of Agriculture, Agricultural Research Service Culture Collection (NRRL), Patent Collection

	NRRL 18488		
N/A	*Pseudomonas putida* KT2440 (ATCC 47054)	Wild type strain	American Type Culture Collection (ATCC)
JPUB_013737	JBEI-107002	*E. coli* DH5α containing pet28a-TcsA	This study
JPUB_013739	JBEI-107003	*E. coli* DH5α containing pet28a-TcsA S98A	This study
JPUB_013727	JBEI-106997	*E. coli* NEB Turbo containing pet28a-TcsD	This study
JPUB_013729	JBEI-106998	*E. coli* DH5α containing pet28a-TcsD L83A	This study
JPUB_013731	JBEI-106999	*E. coli* DH5α containing pet28a-TcsD F79A	This study
JPUB_013733	JBEI-107000	*E. coli* DH5α containing pet28a-TcsD I363A	This study
JPUB_013735	JBEI-107001	*E. coli* DH5α containing pet28a-TcsD F79AL83A	This study
JPUB_013743	JBEI-107005	*E. coli* DH5α containing pet28a-TcsD F79AI363A	This study

JPUB_013741	JBEI-107004	*E. coli* DH5α containing pBbE7a-PP2216-6xHis	This study
JPUB_013841	JBEI-133594	*E. coli* DH5α containing pBbE7a-NmexACP-6xHis	This study
JPUB_013843	JBEI-133595	*E. coli* DH5α containing pBbE7a-NmexACAD-6xHis	This study

Table 3-3 Primers used in this study

Primer name	Sequence (5′→ 3′)
TcsA-NdeI-F	TATACATATGGCGTTCCTCTTCCCCGGC
TcsA-XhoI-R	TATACTCGAGCGCCGCCCGGAAACGGAA
TcsD-NdeI-F	AAACATATGAGCGAATCCGAACGCCTCGGTA
TcsD-XhoI-R	TTTCTCGAGGGTACGTTTCGCGGTGGGGACG
TcsA-AT0-F-1	tcgccgggcacgctctcgggga
TcsA-AT0-R-1	ccactggtaacaggattagcagagcgaggtatgt
TcsA-AT0-F-2	caccgcctacatacctcgctctgctaatcc
TcsA-AT0-R-2	gctccgaactccccgagagcgtgc
TcsD-mut-F	CTCTGTAGCACCGCCTACATACCTCGC
TcsD-mut-R	ACTGGTAACAGGATTAGCAGAGCGAGG
TcsD-F79A-F	GGTGGCGGCCACCCTGTTTCTGC
TcsD-F79A-R	AGGACGGGCAGAAACAGGGTGGCCGCC
TcsD-L83A-F	accctgtttgcgcccgtcctgac

TcsD-L83A-R	gctggtcgtcaggacgggcgcaaacagggt
TcsD-I363-bb-F	GACCCAGAGCGCTGCCGGCACCT
TcsD-I363-bb-R	ctcgtaggacaggtgccggcagcgc
TcsD-I363A-F	cacgcttcgatcgccgagggcggcgacgac
TcsD-I363A-R	gtcgtcgccgccctcggcgatcgaagcgtg
PP2216-E7a-F	CAAAAGATCTTTTAAGAAGGAGATATACATATGCTGGTAAA TGACGAGCAACAAC
PP2216-E7a-R	CTCGAGCCGCCGCCCGGAAACGGAACCGGTAAGATTGCGCG CAATGACCATGC
E7a-link-His-F	GGATCTGGCACTGGTAGT
E7a-link-His-R	CATATGTATATCTCCTTCTTAAAAGATCT
E7a-link-GG-F	CAGGTCTCAGGATCTGGCACTGGTAGT
BioB-GG-R	CAGGTCTCACATATGTATATCTCCTTCTTAAAAGATCTT

Table 3-4 Synthetic DNA fragments used in this study

DNA fragment name	Sequence (5′-->3′)
Nocardia mexicana ACP	caggtctcatATGAACGCGATAACTACAACCGAGATCGCCGAGGGCCT GCGGAGCATCGCCGATCGTCTGGACCTGGAACTCGAGAATGTCG ATATTTCCGCGACGTCGTCGCTGGAAGACGACCTCGAGATGGACT CGCTGAACCTGATGGACTTCCTGGTGTATCTCGAGAAGAGATATC ACGTGCAGGTCACCGACGAGCGCCTGCACGAGGTCGACACCATC GGCGATGTCGTGAACCTGCTCAACGACCTGCTGGCATCGTCGCGC GACACCCCGCAGCCGTCCCGAGTCGGAAGTCGTgg atatgagacctg
Nocardia mexicana ACAD	caggtctcatATGGGCGACATTGTCGAGCAGGCCCGCCATTTTGGACGT ACAGTTCTAGCTGCTGCCCCAGCACCTGATGTAGAAACCGTCTAC GACGACCGGCACCCACTGTGGGAGCAGTTCCGATCGGCCGACCT CGCCGACTGGTGGGTGCCCGCCGAATACGGCGGACGGGGTGTCG GACTGTGCGAGTCGGTGAACGTCGTCTCCGAACTCTCCTACCACG ACGCCGGATTCGCGTTCGCCGCGTTTCTGCCGATCCTCGCGTCACG AATGCTCGAGCTCTACGGTCCCGAGGAGTTGGCGCGCCGCTACTT GGCGGAGATGGCCACCCATGGATCCTTTGCAGCAGCGTTGGGAA GTGAAGCTGAAGCTGGAAGTGAATTAGCTAGAACGCAGACCACG TTCCGCCGCGACGGCGATGTGCTGCACATCAACGGCGATAAGCA GTTCTCGACCAACTTGGCGTTCGCGCGGTTCTGCCTCGTGCTGGCC CGCGACGTCGACAACCCCCGAGATTTCGCCCTGATCCTGGTCCCG TCGGACAGTCCGGGATTTGTTGTCGGCCAACGGTGGCCGATGTCG GGCCTGCACGGTACTGCCACGTACCCGGCAACTTTCACCGATTGC GTCGTCCCGGCAGCAAATCGGTTGTCGGGCAACGGTATTCGGATC CTGGAGGTCGGGCTCGACGCGAGCCGGATCCTGATGGCCTCGATC GCGATCGGTCTGACCCGGCGGATTCGGGATCTGAGCATGGACTAC GCGGCGAGCAAGCGGCTCGGTGGGCAACCGCTGAACCGCAACGC CGTCTTCGGCGCCCGCATGGGTCAGCTCGAGATGGAACTCGAAAC GATGAAGGCCCAGTGCCGGTGCGCCGCGGCCGAATACGACGACA TCTACCAGCGATCCGATCGGGCGGCGGTGTTCTATGCCGACGGCG TGCTGAAGTCGGCGATTGTGGCCAAGATGCATTGTGGTCAGGTCG GCTGGCGGGTCGCGAGTCGGGCGTCGGAGGCGTTCGGCGGTCTGG GCTACACCGCGGGCCACGACATCCAGCGGTGCCTGCGGGACATG CGGCACATCGCGATCGTCGAGGGCGGCGACGACGTGCTGCGTGA ACTGCTCTACGGACGCTACGTCAAGCGCGCATCGCGACGAGGAgg atatgagacctg

3.5.6. Protein expression

All proteins were expressed in *E. coli* BL21 (DE3) cells using the following protocol. An overnight culture of Luria Broth containing the appropriate antibiotic was used to inoculate 0.8 L of Terrific Broth (EMD Millipore) in a 2 L flask. Cells were grown at 37°C to an OD of 0.6-0.8 then induced with 250 µM IPTG (Teknova, USA). Cells were grown for approximately 18 hours at 18°C then pelleted and stored at -20°C until purification.

3.5.7. Protein Purification for biochemical assays and substrate biosynthesis

Frozen cell pellets were resuspended in lysis buffer that had been prechilled to 4°C. For TcsD, TcsD mutants, and the *Nocardia mexicana* ACAD, lysis buffer consisted of 50 mM Tris pH 9.0, 300 mM NaCl, 10 mM imidazole, 8% glycerol. For all other proteins (TcsAS98A, Sfp, PP2216, MatB T207G/M306I, *Nocardia mexicana* ACP), a lysis buffer composed of 50 mM sodium phosphate pH 7.2, 300 mM NaCl, 10 mM imidazole, 8% glycerol was used. After resuspension in lysis buffer, cells were sonicated on ice to lyse. Lysates were then centrifuged at 4°C at 40,000g for 30 minutes to pellet insolubles. The supernatants were then applied to Ni-NTA resin in an Econo-Pak column (Bio-Rad, USA) that had been pre-equilibrated with 10 column volumes of lysis buffer. The resin was washed with several column volumes of lysis buffer and subsequently with several column volumes of lysis buffer containing 50 mM imidazole. The proteins were then eluted using a stepwise gradient of lysis buffer containing 100 mM, 200 mM and 400 mM imidazole. Fractions were analyzed via SDS-PAGE, and fractions containing >90% pure protein were pooled and concentrated using Amicon Ultra 100 kDa (TcsD, TcsDmutants, and *Nocardia mexicana* ACAD), 3kDa (*Nocardia mexicana* ACP)10 kDa (all other proteins) Molecular Weight Cutoff (MWCO) centrifuge filters (Millipore). Proteins were transferred to storage buffer (lysis buffer containing no imidazole, 100 mM NaCl, and 8-10% glycerol) by buffer exchanging in the centrifuge filters and frozen in liquid nitrogen. Proteins were stored at -80°C until assays.

3.5.8. Protein purification for crystallography

A frozen cell pellet of *E. coli* BL21 (DE3) expressing TcsD was purified using Ni-NTA affinity chromatography as described above. After concentration to a minimal volume, the protein was further purified using size exclusion chromatography. It was loaded onto a HiPrep 26/60 Sephacryl S-300 High Resolution column that had been pre-equilibrated with size exclusion buffer (50 mM Tris pH 9, 500 mM NaCl, 10%

glycerol) at 4°C. The protein was eluted and concentrated again then dialyzed overnight at 4°C into crystallization buffer (50 mM Tris pH 9, 100 mM NaCl).

3.5.9. Protein Crystallization and Structure Determination

An initial crystallization screen was set up using a Phoenix robot (Art Robbins Instruments, Sunnyvale, CA) using the sparse matrix screening method.[268] The purified TcsD protein sample was concentrated to 9 mg/mL and crystallized at 25°C using the sitting drop method in 0.4 µL drops containing a 1:1 ratio of protein to crystallization solution: 0.1 M Sodium Citrate tribasic dihydrate, pH 5.0, and 10% PEG 6,000. Crystals were transferred to crystallization solution containing 20% glycerol prior to flash freezing in liquid nitrogen. X-ray diffraction data was collected at the Berkeley Center for Structural Biology on beamline 5.0.2 of the Advanced Light Source at the Lawrence Berkeley National Laboratory. The TcsD structure was determined by the molecular-replacement method with the program *PHASER*[269] using medium-chain acyl-CoA dehydrogenase from *Thermus thermophilus* (PDB ID: 1UKW) as the search model, which showed 32% sequence identity. Structure refinement was performed by *phenix.refine* program.[270] Manual rebuilding using COOT[271] and the addition of water molecules allowed for construction of the final model. The final models of the TcsD structure showed an R-work of 14.6% and R-free of 17.9%. Root-mean-square deviations from ideal geometries for bond lengths, angles, and dihedrals were calculated with Phenix.[272] The overall stereochemical quality of the final models for tcsD was assessed using the MolProbity program.[273] The structural analysis was performed in COOT and PyMOL.[274]

Table 3-5 Summary of crystal parameters, data collection, and refinement statistics. Values in parentheses are for the highest resolution shell.

Crystal parameters	
Space group	P 1 21 1
Unit cell	69.52 100.06 148.51 90 100.31 90
Data collection statistics	
Wavelength (Å)	1.00000
Resolution range (Å)	68.4 - 1.75 (1.813 - 1.75)
Total reflections	835542 (85389)
Unique reflections	196484 (19281)
Multiplicity	4.3 (4.4)
Completeness (%)	97.65 (96.41)
Mean I/sigma(I)	8.99 (1.39)

Wilson B-factor	19.7
R-merge	0.1119 (1.063)
R-meas	0.1277 (1.207)
R-pim	0.06086 (0.5664)
CC1/2	0.997 (0.524)
CC*	0.999 (0.829)
Refinement and model statistics	
Reflections used in refinement	196470 (19281)
Reflections used for R-free	9864 (1003)
R-work	0.1462 (0.2810)
R-free	0.1787 (0.3138)
CC(work)	0.974 (0.807)
CC(free)	0.959 (0.765)
Number of non-hydrogen atoms	13452
macromolecules	11781
ligands	212
solvent	1459
Protein residues	1546
RMS(bonds)	0.012
RMS(angles)	0.98
Ramachandran favored (%)	99.28
Ramachandran allowed (%)	0.72
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.08
Clashscore	3.04
Average B-factor	27.08
macromolecules	26.03
ligands	17.5
solvent	36.93
Number of TLS groups	31

3.5.10. Synthesis of valeroyl-Coenzyme A

Valeroyl-Coenzyme A was prepared using the anhydride method.[275] Approximately 0.02 mmol of Coenzyme A was added to a solution of saturated sodium bicarbonate in water. The solution was chilled to 0°C, then 1 mmol of valeric anhydride

(Sigma Aldrich) was added. The reaction was allowed to proceed at 0°C with constant mixing. After approximately 6 hours, HCl was added to bring the pH to ~2. The solution was extracted twice with an equal volume of ethyl acetate, then the remaining aqueous solution was frozen until further purification.

Valeroyl-CoA was purified using an Agilent 1260 series preparatory HPLC system equipped with a 900µL sample loop and a UV detector coupled to a fraction collector. The product was isolated over an Agilent Prep-C18 column (21.2 x 150 mm, 5 µm pore size) using a mobile phase composed of 10 mM ammonium formate, pH 4.5 (solvent A) and methanol (solvent B) using the following method:

Table 3-6 Preparatory LC gradient for purification of valeroyl-Coenzyme A

Time (min)	%A	%B	Flow rate (mL/min)
0	95	5	10
1	95	5	10
8	5	95	10
10	5	95	10
11	95	5	10
16	95	5	10

The fraction collector was programmed to collect all peaks that absorbed at 280 nm. Fractions containing valeroyl-CoA were identified by direct infusion into an Applied Biosystems 4000 QTRAP mass spectrometer with the following mass spectrometer parameters: Q1 MS mode, scan range 100-1000 m/z, declustering potential: 70, entrance potential: 10, curtain gas: 10, IonSpray Voltage: 4800, Temperature: 300, Ion Source Gases: 40. The fractions containing the correct m/z for valeroyl-Coenzyme A ([M+H]) were then combined and lyophilized, leaving behind valeroyl-CoA as a white powder. The product was resuspended in LC grade water and quantified using the absorbance at 280 nm compared to a standard curve of hexanoyl-CoA in water (Sigma-Aldrich) measured on a Nanodrop ND-1000 Spectrophotometer. The identity of the Coenzyme A ester was additionally verified by targeted LC-MS/MS after loading the purified substrate onto TcsA S98A (methods described below).

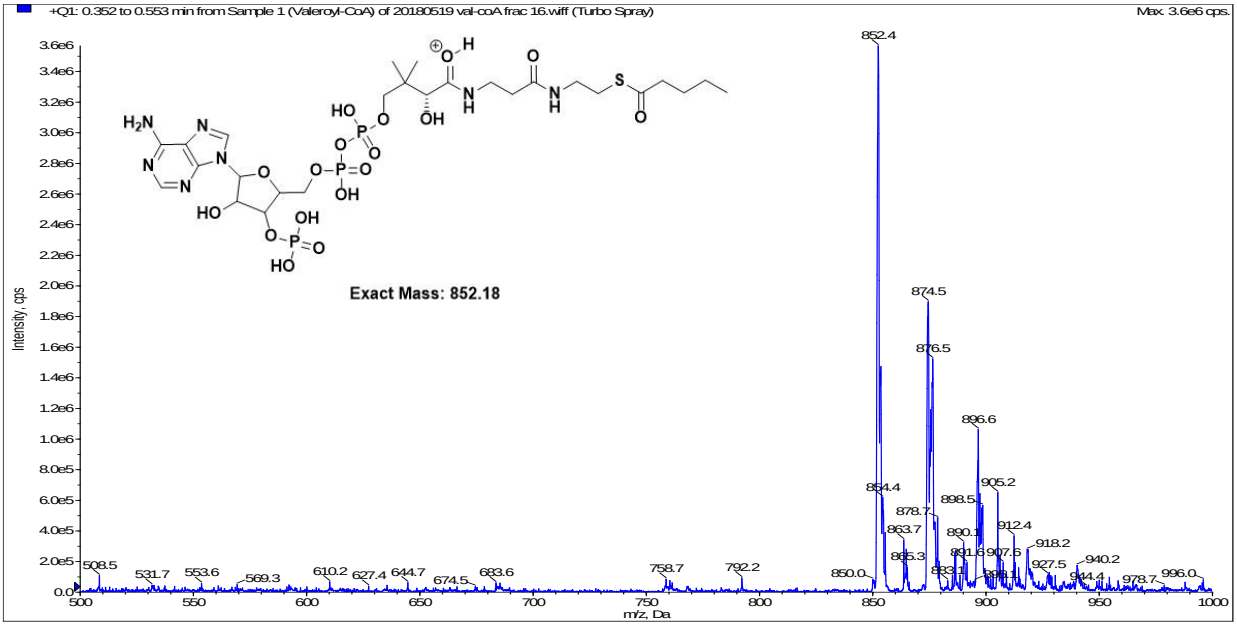


Figure 3-6 Mass spectrum of synthesized valeroyl-CoA. An m/z of 852.18 [z=1] represents the [M+H] ion, while the m/z of 874.5 [z=1] represents the [M+Na] ion.

3.5.11. Trans-2-pentenoyl-, trans-2-hexenoyl-, and trans-2-heptenoyl-Coenzyme A.

The 2,3-unsaturated CoA substrates were generated *in vitro* from valeroyl-CoA (synthesized in this work), hexanoyl-CoA (Sigma Aldrich), or heptanoyl-CoA (CoALA Biosciences, USA) using the short chain acyl-CoA dehydrogenase PP2216.[261] Reactions contained the following components: 100mM sodium phosphate (pH 7.2), 500 µM Acyl-Coenzyme A, 1 mM ferrocenium hexafluorophosphate (FeHFP, Sigma Aldrich), 225 µM PP2216 in a final volume of 1 mL. FeHFP was always prepared fresh by dissolving solid FeHFP in 10mM HCl and sonicating in a water bath to dissolve solid particles.[254] Reactions were incubated at 25°C for 2 hours (2-pentenoyl-CoA) or 6 hours (2-hexenoyl and 2-heptenoyl-CoA) then filtered through 3kDa MWCO Amicon microcentrifuge filters (Sigma Aldrich) to remove the remaining PP2216 protein. Samples were then aliquoted and lyophilized. After lyophilization, samples were stored at -20°C and resuspended in HPLC water (Honeywell) at a concentration of 4mM (assuming maximum theoretical yield) immediately prior to use.

3.5.12. Allylmalonyl- and propylmalonyl-Coenzyme A.

Allylmalonyl-CoA and propylmalonyl-CoA were biosynthesized from the corresponding malonic acids using an engineered mutant of the *Streptomyces coelicolor* malonyl-Coenzyme A transacylase MatB, MatB T207G/M306I.[150] Reactions contained the following components: 100 mM sodium phosphate (pH 7.2), 2 mM $MgCl_2$,

8 mM ATP, 4 mM Coenzyme A, 4 mM diacid, 100 µM MatB T207G/M306I in a final volume of 0.5 mL. Prior to addition to assays, allylmalonic acid (Sigma Aldrich) and propylmalonic acid (VWR, BeanTown Chemical) were dissolved in 50 mM sodium phosphate buffer, pH 7.2. Reactions were incubated for 3 hours the filtered through 3kDa MWCO Amicon microcentrifuge filters (Sigma Aldrich) to remove the remaining MatB protein. Mixtures were aliquoted and frozen at -20°C until use in TcsD assays.

3.5.13. Enzyme Assays: TcsD and mutants

Assays of TcsD activity on TcsAS98A-bound substrates were carried out as follows: First, purified TcsAS98A was loaded with acyl-Coenzyme A substrates via the action of the promiscuous phosphopantetheinyl transferase Sfp.[166] Loading reactions (150 µL) contained 60 µM TcsAS98A, 10 µM Sfp, 250 µM acyl-Coenzyme A, 20 mM $MgCl_2$, and 50 mM sodium phosphate, pH 7.2. Reactions were initiated by the addition of Sfp and allowed to proceed at 37°C for 30 minutes. The loaded protein was then immediately used for TcsD assays, each of which contained: 45 µL of loading reaction, 0.5 mM FeHFP (dissolved in 10 mM HCl), and 2 µM TcsD mutant (or boiled TcsD mutant). FeHFP was prepared as described above. Reactions were allowed to proceed for 16 hours then quenched with and equal volume of HPLC methanol (Sigma Aldrich) and either stored at -20°C or immediately processed for LC-MS analysis.

3.5.14. Enzyme assays: *Nocardia mexicana* ACAD

The *Nocardia mexicana* ACP was loaded with butyryl-, pentanoyl-, 2-pentenoyl-, 2-hexenoyl, or 2-heptenoyl-CoA using Sfp.[166] Loading reactions (120 µL) consisted of 50 µM *N. mexicana* ACP, 10 µM Sfp, 250 µM acyl-Coenzyme A, 20 mM $MgCl_2$, and 40 mM Tris buffer, pH 8. Reactions were initiated by the addition of Sfp and allowed to proceed at 37°C for 1 hour. The loading reactions were then used for *N. mexicana* ACAD assays which consisted of: 40 µL of loading reaction, 0.5 mM FeHFP,[254] and 2 µM *N. mexicana* ACAD. Reactions were allowed to proceed at 37°C for 16 hours then quenched with HPLC methanol and immediately processed for LC-MS analysis.

3.5.15. Sample preparation for LC-MS/MS analysis

Samples were prepared for proteomics analysis by chloroform/methanol extraction as previously reported.[276] Briefly, 150 µL of LC-grade methanol and 50 µL of LC-grade chloroform were sequentially added to the methanol-quenched samples, and samples were vortexed after each addition. Then, 150 µL LC-grade water was added, samples were vortexed to mix, and they were subsequently centrifuged at 21,000g for 1 minute to promote phase separation. After removal and disposal of the top (aqueous) layer, another 150 µL of LC methanol was added, samples were vortexed, then they were centrifuged again at 21,000g for 2 minutes. After removal of the supernatant, samples

were allowed to dry for 5 minutes in a fume hood. Then, they were resuspended in freshly-prepared 100 mM ammonium bicarbonate buffer (*N. mexicana* ACAD assays) or 100 mM ammonium bicarbonate containing 20% LC-grade DMSO (TcsD assays). Protein concentrations were analyzed via the DC Assay (Bio-Rad). After quantification, samples were digested with an appropriate protease. Samples digested with only trypsin (Sigma-Aldrich) or only chymotrypsin (Promega) were digested at a ratio of 1:50 w/w protease:protein sample at 37°C for at least 6 hours. For Asp-N/trypsin dual digestions (TcsD and TcsD mutant assays), proteins were digested with 1:25 w/w Asp-N (New England Biolabs) in the presence of 0.75 mM $ZnSO_4$ for 2-4 hours at 37°C then immediately digested again with 1:10 w/w trypsin for 1 hour at 37 °C. The *N. mexicana* ACAD assays were digested with Glu-C (Promega) at a ratio of 1:50 protease:protein sample at 37°C for 6 hours. After digestion, the samples were either frozen for future analysis or directly analyzed via LC-MS/MS.[255]

3.5.16. Shotgun LC-MS/MS Analysis of TcsA-bound substrates

Samples prepared for shotgun proteomic analysis were analyzed using Agilent 6550 iFunnel Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA) coupled to an Agilent 1290 UHPLC system. Twenty (20) µg of peptides were separated on a Sigma–Aldrich Ascentis Peptides ES-C18 column (2.1 mm × 100 mm, 2.7 µm particle size, operated at 60°C) at a 0.400 mL/min flow rate and eluted with the following gradient: initial condition was 95% solvent A (0.1% formic acid) and 5% solvent B (99.9% acetonitrile, 0.1% formic acid). Solvent B was increased to 35% over 5.5 min, and then increased to 80% over 1min, and held for 3.5 min at a flow rate of 0.6 mL/min, followed by a ramp back down to 5% B over 0.5 min where it was held for 2 min to re-equilibrate the column to original conditions. Peptides were introduced to the mass spectrometer from the LC by using a Jet Stream source (Agilent Technologies) operating in positive-ion mode (3,500 V). Source parameters employed gas temp (250°C), drying gas (14 L/min), nebulizer (35 psig), sheath gas temp (250°C), sheath gas flow (11 L/min), VCap (3,500 V), fragmentor (180 V), OCT 1 RF Vpp (750 V). The data were acquired with Agilent MassHunter Workstation Software, LC/MS Data Acquisition B.06.01 operating in Auto MS/MS mode whereby the 20 most intense ions (charge states, 2–5) within 300–1,400 m/z mass range above a threshold of 1,500 counts were selected for MS/MS analysis. MS/MS spectra (100–1,700 m/z) were collected with the quadrupole set to "Medium" resolution and were acquired until 45,000 total counts were collected or for a maximum accumulation time of 333 ms. Former parent ions were excluded for 0.1 min following MS/MS acquisition. The acquired data were exported as mgf files and searched against a customized protein database with Mascot search engine version 2.3.02 (Matrix Science).

3.5.17. Targeted LC-MS/MS Analysis

Samples were analyzed using an Agilent 1290 Infinity II liquid chromatography system coupled to an Agilent 6460 QQQ mass spectrometer (Agilent Technologies, Santa Clara, CA). Peptide samples (10 µg) were separated on an Ascentis Express Peptide ES-C18 column (2.7 µm particle size, 160 Å pore size, 50 x 2.1mm) fitted with a guard column (5 mm x 2.1 mm, Sigma Aldrich). The column was heated to 60°C. The mobile phase consisted of 0.1% formic acid in H_2O (A) and 0.1% formic acid in acetonitrile (B). Peptides were eluted via the following gradient method:

Table 3-7 LC gradient for targeted LC-MS/MS analysis of phosphopantetheine-bearing peptides

Time (min)	%A	%B	Flow rate (mL/min)
0	95	5	0.4
2.2	95	5	0.4
7.7	65	35	0.4
8	20	80	0.4
10	20	80	0.4
10.5	95	5	0.4
12	95	5	0.4
15	95	5	0.4

Peptides were ionized using an Agilent Jet Stream ESI source operating in positive-ion mode with the following source parameters: Gas Temperature = 250°C, Gas Flow = 13 L/min, Nebulizer Pressure = 35 psi, Sheath Gas Temperature = 250°C, Sheath Gas Flow = 11 L/min, and Capillary Voltage = 3,500 V. Dwell times were set to 10ms. Data were acquired using Agilent MassHunter Data Acquisition (Version B.08.02, Build 8.2.8260.0). The mass spectrometry method was built using the Skyline targeted mass spectrometry environment (version 4.2.0 19072).[277] Collision energies were predicted by Skyline with the exception of phosphopantetheine-bearing peptides for which the collision energy was set to 30eV. The specific transitions monitored and the collision energies for each are listed in Tables 3-8 & 3-9.

Table 3-8 Transitions used for the targeted LC-MS/MS-based detection of TcsA-bound intermediates

Protein	Peptide	Precursor m/z	Precursor Charge	Product m/z	Product Charge	Collision Energy (eV)
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	APPPTGGMLAVK	569.818084	2	970.539014	1	18.7
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	APPPTGGMLAVK	569.818084	2	873.48625	1	18.7
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	APPPTGGMLAVK	569.818084	2	430.302396	1	18.7
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	AGELIAAAR	436.253434	2	501.314357	1	14.5
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	AGELIAAAR	436.253434	2	388.230293	1	14.5
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	AGELIAAAR	436.253434	2	317.193179	1	14.5
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	DASALYATTMR	600.289887	2	742.355236	1	19.6
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	DASALYATTMR	600.289887	2	579.291907	1	19.6
tr|E9KTG6|E9KTG6_9ACTN_TcsAS98A	DASALYATTMR	600.289887	2	508.254793	1	19.6
sp|P39135|SFP_BACSU	TKPISLEIAK	550.339706	2	870.529495	1	18.1
sp|P39135|SFP_BACSU	TKPISLEIAK	550.339706	2	573.360639	1	18.1
sp|P39135|SFP_BACSU	TKPISLEIAK	550.339706	2	460.276575	1	18.1
sp|P39135|SFP_BACSU	TKPISLEIAK	550.339706	2	331.233982	1	18.1
Holo-ACP	DLGVDSLAMTELQAHALQR[+340.085794]	1204.571117	2	261.126739	1	30
Holo-ACP	DLGVDSLAMTELQAHALQR[+340.085794]	803.38317	3	261.126739	1	30
Holo-ACP	DSLAMTELQAHALQR[+340.085794]	1012.470675	2	261.126739	1	30
Holo-ACP	DSLAMTELQAHALQR[+340.085794]	675.316209	3	261.126739	1	30
Apo-ACP	DLGVDSLAMTELQAHALQR	1034.52822	2	1297.668131	1	33.1
Apo-ACP	DLGVDSLAMTELQAHALQR	1034.52822	2	1166.627646	1	33.1

Apo-ACP	DLGVDSLAMTELQAHALQR	1034.52822	2	1065.579967	1	33.1
Apo-ACP	DLGVDSLAMTELQAHALQR	1034.52822	2	936.537374	1	33.1
Apo-ACP	DLGVDSLAMTELQAHALQR	1034.52822	2	823.45331	1	33.1
Apo-ACP	DLGVDSLAMTELQAHALQR	690.021239	3	936.537374	1	20
Apo-ACP	DLGVDSLAMTELQAHALQR	690.021239	3	823.45331	1	20
Apo-ACP	DLGVDSLAMTELQAHALQR	690.021239	3	695.394733	1	20
Apo-ACP	DLGVDSLAMTELQAHALQR	690.021239	3	624.357619	1	20
Apo-ACP	DLGVDSLAMTELQAHALQR	690.021239	3	487.298707	1	20
Apo-ACP	DSLAMTELQAHALQR	842.427778	2	1166.627646	1	27.1
Apo-ACP	DSLAMTELQAHALQR	842.427778	2	1065.579967	1	27.1
Apo-ACP	DSLAMTELQAHALQR	842.427778	2	936.537374	1	27.1
Apo-ACP	DSLAMTELQAHALQR	842.427778	2	823.45331	1	27.1
Apo-ACP	DSLAMTELQAHALQR	842.427778	2	695.394733	1	27.1
Apo-ACP	DSLAMTELQAHALQR	561.954277	3	823.45331	1	15.4
Apo-ACP	DSLAMTELQAHALQR	561.954277	3	695.394733	1	15.4
Apo-ACP	DSLAMTELQAHALQR	561.954277	3	624.357619	1	15.4
Apo-ACP	DSLAMTELQAHALQR	561.954277	3	487.298707	1	15.4
Apo-ACP	DSLAMTELQAHALQR	561.954277	3	416.261593	1	15.4
2-pentenoyl-ACP	DLGVDSLAMTELQAHALQR[+422.127659]	1245.59205	2	343.168604	1	30
2-pentenoyl-ACP	DLGVDSLAMTELQAHALQR[+422.127659]	830.730459	3	343.168604	1	30
2-pentenoyl-ACP	DSLAMTELQAHALQR[+422.127659]	1053.491607	2	343.168604	1	30
2-pentenoyl-ACP	DSLAMTELQAHALQR[+422.127659]	702.663497	3	343.168604	1	30

2_4-pentadienoyl-ACP	DLGVDSLAMTELQAHALQR[+420.112009]	1244.584225	2	341.152954	1	30
2_4-pentadienoyl-ACP	DLGVDSLAMTELQAHALQR[+420.112009]	830.058575	3	341.152954	1	30
2_4-pentadienoyl-ACP	DSLAMTELQAHALQR[+420.112009]	1052.483782	2	341.152954	1	30
2_4-pentadienoyl-ACP	DSLAMTELQAHALQR[+420.112009]	701.991614	3	341.152954	1	30
Hydroxypentanoyl-ACP	DLGVDSLAMTELQAHALQR[+440.138224]	1254.597332	2	361.179169	1	30
Hydroxypentanoyl-ACP	DLGVDSLAMTELQAHALQR[+440.138224]	836.73398	3	361.179169	1	30
Hydroxypentanoyl-ACP	DSLAMTELQAHALQR[+440.138224]	1062.49689	2	361.179169	1	30
Hydroxypentanoyl-ACP	DSLAMTELQAHALQR[+440.138224]	708.667019	3	361.179169	1	30
Dihydroxypentanoyl-ACP	DLGVDSLAMTELQAHALQR[+456.133138]	1262.594789	2	377.174083	1	30
Dihydroxypentanoyl-ACP	DLGVDSLAMTELQAHALQR[+456.133138]	842.065618	3	377.174083	1	30
Dihydroxypentanoyl-ACP	DSLAMTELQAHALQR[+456.133138]	1070.494347	2	377.174083	1	30
Dihydroxypentanoyl-ACP	DSLAMTELQAHALQR[+456.133138]	713.998657	3	377.174083	1	30
3-hydroxy-45-pentenoyl-ACP	DLGVDSLAMTELQAHALQR[+438.122574]	1253.589507	2	359.163519	1	30
3-hydroxy-45-pentenoyl-ACP	DLGVDSLAMTELQAHALQR[+438.122574]	836.062097	3	359.163519	1	30
3-hydroxy-45-pentenoyl-ACP	DSLAMTELQAHALQR[+438.122574]	1061.489065	2	359.163519	1	30
3-hydroxy-45-pentenoyl-ACP	DSLAMTELQAHALQR[+438.122574]	707.995135	3	359.163519	1	30
2-hexenoyl-ACP	DLGVDSLAMTELQAHALQR[+436.143309]	1252.599875	2	357.184254	1	30
2-hexenoyl-ACP	DLGVDSLAMTELQAHALQR[+436.143309]	835.402342	3	357.184254	1	30
2-hexenoyl-ACP	DSLAMTELQAHALQR[+436.143309]	1060.499432	2	357.184254	1	30
2-hexenoyl-ACP	DSLAMTELQAHALQR[+436.143309]	707.335 38	3	357.184254	1	30
Hydroxyhexanoyl-ACP	DLGVDSLAMTELQAHALQR[+454.153874]	1261.605157	2	375.194819	1	30
Hydroxyhexanoyl-ACP	DLGVDSLAMTELQAHALQR[+454.153874]	841.405864	3	375.194819	1	30

Hydroxyhexanoyl-ACP	DSLAMTELQAHALQR[+454.153874]	1069.504715	2	375.194819	1	30
Hydroxyhexanoyl-ACP	DSLAMTELQAHALQR[+454.153874]	713.338902	3	375.194819	1	30
3-hydroxy-45-hexenoyl-ACP	DLGVDSLAMTELQAHALQR[+452.138224]	1260.597332	2	373.179169	1	30
3-hydroxy-45-hexenoyl-ACP	DLGVDSLAMTELQAHALQR[+452.138224]	840.73398	3	373.179169	1	30
3-hydroxy-45-hexenoyl-ACP	DSLAMTELQAHALQR[+452.138224]	1068.49689	2	373.179169	1	30
3-hydroxy-45-hexenoyl-ACP	DSLAMTELQAHALQR[+452.138224]	712.667019	3	373.179169	1	30
Dihydroxyhexanoyl-ACP	DLGVDSLAMTELQAHALQR[+470.148788]	1269.602614	2	391.189733	1	30
Dihydroxyhexanoyl-ACP	DLGVDSLAMTELQAHALQR[+470.148788]	846.737502	3	391.189733	1	30
Dihydroxyhexanoyl-ACP	DSLAMTELQAHALQR[+470.148788]	1077.502172	2	391.189733	1	30
Dihydroxyhexanoyl-ACP	DSLAMTELQAHALQR[+470.148788]	718.67054	3	391.189733	1	30
2_4-hexadienoyl-ACP	DLGVDSLAMTELQAHALQR[+434.127659]	1251.59205	2	355.168604	1	30
2_4-hexadienoyl-ACP	DLGVDSLAMTELQAHALQR[+434.127659]	834.730459	3	355.168604	1	30
2_4-hexadienoyl-ACP	DSLAMTELQAHALQR[+434.127659]	1059.491607	2	355.168604	1	30
2_4-hexadienoyl-ACP	DSLAMTELQAHALQR[+434.127659]	706.663497	3	355.168604	1	30
tr|E9KTG9|E9KTG9_9ACTN_TcsD	DAPLALYER	524.277106	2	861.482879	1	17.3
tr|E9KTG9|E9KTG9_9ACTN_TcsD	DAPLALYER	524.277106	2	580.308937	1	17.3
tr|E9KTG9|E9KTG9_9ACTN_TcsD	DAPLALYER	524.277106	2	467.224873	1	17.3
tr|E9KTG9|E9KTG9_9ACTN_TcsD	DAPLALYER	524.277106	2	304.161545	1	17.3
tr|E9KTG9|E9KTG9_9ACTN_TcsD	YTAVTVPR	453.755809	2	472.287808	1	15.1
tr|E9KTG9|E9KTG9_9ACTN_TcsD	YTAVTVPR	453.755809	2	371.24013	1	15.1
tr|E9KTG9|E9KTG9_9ACTN_TcsD	YTAVTVPR	453.755809	2	272.171716	1	15.1
2-heptenoyl-ACP	DLGVDSLAMTELQAHALQR[+450.158959]	1259.6077	2	371.199904	1	30

2-heptenoyl-ACP	DLGVDSLAMTELQAHALQR[+450.158959]	840.074225	3	371.199904	1	30
2-heptenoyl-ACP	DSLAMTELQAHALQR[+450.158959]	1067.507257	2	371.199904	1	30
2-heptenoyl-ACP	DSLAMTELQAHALQR[+450.158959]	712.007264	3	371.199904	1	30
2_4-heptadienoyl-ACP	DLGVDSLAMTELQAHALQR[+448.143309]	1258.599875	2	369.184254	1	30
2_4-heptadienoyl-ACP	DLGVDSLAMTELQAHALQR[+448.143309]	839.402342	3	369.184254	1	30
2_4-heptadienoyl-ACP	DSLAMTELQAHALQR[+448.143309]	1066.499432	2	369.184254	1	30
2_4-heptadienoyl-ACP	DSLAMTELQAHALQR[+448.143309]	711.33538	3	369.184254	1	30
Hydroxyheptanoyl-ACP	DLGVDSLAMTELQAHALQR[+468.169524]	1268.612982	2	389.210469	1	30
Hydroxyheptanoyl-ACP	DLGVDSLAMTELQAHALQR[+468.169524]	846.077747	3	389.210469	1	30
Hydroxyheptanoyl-ACP	DSLAMTELQAHALQR[+468.169524]	1076.51254	2	389.210469	1	30
Hydroxyheptanoyl-ACP	DSLAMTELQAHALQR[+468.169524]	718.010785	3	389.210469	1	30
Dihydroxyheptanoyl-ACP	DLGVDSLAMTELQAHALQR[+484.164439]	1276.61044	2	405.205384	1	30
Dihydroxyheptanoyl-ACP	DLGVDSLAMTELQAHALQR[+484.164439]	851.409385	3	405.205384	1	30
Dihydroxyheptanoyl-ACP	DSLAMTELQAHALQR[+484.164439]	1084.509997	2	405.205384	1	30
Dihydroxyheptanoyl-ACP	DSLAMTELQAHALQR[+484.164439]	723.342424	3	405.205384	1	30
3-hydroxy-45-heptenoyl-ACP	DLGVDSLAMTELQAHALQR[+466.153874]	1267.605157	2	387.194819	1	30
3-hydroxy-45-heptenoyl-ACP	DLGVDSLAMTELQAHALQR[+466.153874]	845.405864	3	387.194819	1	30
3-hydroxy-45-heptenoyl-ACP	DSLAMTELQAHALQR[+466.153874]	1075.504715	2	387.194819	1	30
3-hydroxy-45-heptenoyl-ACP	DSLAMTELQAHALQR[+466.153874]	717.338902	3	387.194819	1	30
Pentanoyl-ACP	DLGVDSLAMTELQAHALQR[+424.1]	1246.599875	2	345.184254	1	30
Pentanoyl-ACP	DLGVDSLAMTELQAHALQR[+424.1]	831.402342	3	345.184254	1	30
Pentanoyl-ACP	DSLAMTELQAHALQR[+424.1]	1054.499432	2	345.184254	1	30

Pentanoyl-ACP	DSLAMTELQAHALQR[+424.1]	703.33538	3	345.184254	1	30
Butyryl-ACP	DLGVDSLAMTELQAHALQR[+410.1]	1239.59205	2	331.168604	1	30
Butyryl-ACP	DLGVDSLAMTELQAHALQR[+410.1]	826.730459	3	331.168604	1	30
Butyryl-ACP	DSLAMTELQAHALQR[+410.1]	1047.491607	2	331.168604	1	30
Butyryl-ACP	DSLAMTELQAHALQR[+410.1]	698.663497	3	331.168604	1	30
Crotonyl-ACP	DLGVDSLAMTELQAHALQR[+408.1]	1238.584225	2	329.152954	1	30
Crotonyl-ACP	DLGVDSLAMTELQAHALQR[+408.1]	826.058575	3	329.152954	1	30
Crotonyl-ACP	DSLAMTELQAHALQR[+408.1]	1046.483782	2	329.152954	1	30
Crotonyl-ACP	DSLAMTELQAHALQR[+408.1]	697.991614	3	329.152954	1	30
Hydroxybutyryl-ACP	DLGVDSLAMTELQAHALQR[+426.1]	1247.589507	2	347.163519	1	30
Hydroxybutyryl-ACP	DLGVDSLAMTELQAHALQR[+426.1]	832.062097	3	347.163519	1	30
Hydroxybutyryl-ACP	DSLAMTELQAHALQR[+426.1]	1055.489065	2	347.163519	1	30
Hydroxybutyryl-ACP	DSLAMTELQAHALQR[+426.1]	703.995135	3	347.163519	1	30
Propylmalonyl-ACP	DLGVDSLAMTELQAHALQR[+468.1]	1268.594789	2	389.174083	1	30
Propylmalonyl-ACP	DLGVDSLAMTELQAHALQR[+468.1]	846.065618	3	389.174083	1	30
Propylmalonyl-ACP	DSLAMTELQAHALQR[+468.1]	1076.494347	2	389.174083	1	30
Propylmalonyl-ACP	DSLAMTELQAHALQR[+468.1]	717.998657	3	389.174083	1	30
Allymalonyl-ACP	DLGVDSLAMTELQAHALQR[+466.1]	1267.586964	2	387.158433	1	30
Allymalonyl-ACP	DLGVDSLAMTELQAHALQR[+466.1]	845.393735	3	387.158433	1	30
Allymalonyl-ACP	DSLAMTELQAHALQR[+466.1]	1075.486522	2	387.158433	1	30
Allymalonyl-ACP	DSLAMTELQAHALQR[+466.1]	717.326773	3	387.158433	1	30

Note: peptides with added masses in brackets (e.g. "DSLAMTELQAHALQR[+466.1]") depict modified peptides with phosphopantetheine arms/substrates.

Table 3-9 Transitions used for the targeted LC-MS/MS-based detection of NmexACP-bound intermediates

Protein	Peptide	Precursor Ion	Precursor charge	Product Ion	Product charge	Collision Energy (eV)
Holo-Nmex-ACP	MDSLNLMDFLVYLE[+340.085794].light	1021.94986	2	261.12674	1	30
Holo-Nmex-ACP	MDSLNLMDFLVYLE[+340.085794].light	681.635668	3	261.12674	1	30
Nmex-ACP	GLRSIADRLDLE.light	679.37534	2	760.38356	1	22.1
Nmex-ACP	GLRSIADRLDLE.light	679.37534	2	645.35662	1	22.1
Nmex-ACP	GLRSIADRLDLE.light	679.37534	2	489.25551	1	22.1
Nmex-ACP	GLRSIADRLDLE.light	453.252652	3	489.25551	1	11.5
Nmex-ACP	GLRSIADRLDLE.light	453.252652	3	376.17144	1	11.5
Nmex-ACP	GLRSIADRLDLE.light	453.252652	3	261.1445	1	11.5
Nmex-ACP	NVDISATSSLEDDLE.light	804.367772	2	820.35707	1	25.9
Nmex-ACP	NVDISATSSLEDDLE.light	804.367772	2	733.32504	1	25.9
Nmex-ACP	NVDISATSSLEDDLE.light	804.367772	2	620.24098	1	25.9
Pentanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+424.143309].light	1063.97862	2	345.18425	1	30
Pentanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+424.143309].light	709.65484	3	345.18425	1	30
Pentenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+422.127659].light	1062.9708	2	343.1686	1	30
Pentenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+422.127659].light	708.982956	3	343.1686	1	30
Pentadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+420.112009].light	1061.96297	2	341.15295	1	30
Pentadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+420.112009].light	708.311073	3	341.15295	1	30
Hydroxypentanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+440.138224].light	1071.97608	2	361.17917	1	30
Hydroxypentanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+440.138224].light	714.986478	3	361.17917	1	30

Butyryl-Nmex-ACP	MDSLNLMDFLVYLE[+410.127659].light	1056.9708	2	331.1686	1	30
Butyryl-Nmex-ACP	MDSLNLMDFLVYLE[+410.127659].light	704.982956	3	331.1686	1	30
Crotonyl-Nmex-ACP	MDSLNLMDFLVYLE[+408.112009].light	1055.96297	2	329.15295	1	30
Crotonyl-Nmex-ACP	MDSLNLMDFLVYLE[+408.112009].light	704.311073	3	329.15295	1	30
Hydroxybutyryl-Nmex-ACP	MDSLNLMDFLVYLE[+426.122574].light	1064.96825	2	347.16352	1	30
Hydroxybutyryl-Nmex-ACP	MDSLNLMDFLVYLE[+426.122574].light	710.314595	3	347.16352	1	30
Hexanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+438.158959].light	1070.98645	2	359.1999	1	30
Hexanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+438.158959].light	714.326723	3	359.1999	1	30
Hexenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+436.143309].light	1069.97862	2	357.18425	1	30
Hexenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+436.143309].light	713.65484	3	357.18425	1	30
Hexadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+434.127659].light	1068.9708	2	355.1686	1	30
Hexadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+434.127659].light	712.982956	3	355.1686	1	30
Hydroxyhexanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+454.153874].light	1078.9839	2	375.19482	1	30
Hydroxyhexanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+454.153874].light	719.658361	3	375.19482	1	30
Heptanoyl-ACP	MDSLNLMDFLVYLE[+452.174609].light	1077.99427	2	373.21555	1	30
Heptanoyl-ACP	MDSLNLMDFLVYLE[+452.174609].light	718.998606	3	373.21555	1	30
Hydroxyheptanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+468.169524].light	1085.99173	2	389.21047	1	30
Hydroxyheptanoyl-Nmex-ACP	MDSLNLMDFLVYLE[+468.169524].light	724.330245	3	389.21047	1	30
Heptenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+450.158959].light	1076.98645	2	371.1999	1	30
Heptenoyl-Nmex-ACP	MDSLNLMDFLVYLE[+450.158959].light	718.326723	3	371.1999	1	30
Heptadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+448.143309].light	1075.97862	2	369.18425	1	30
Heptadienoyl-Nmex-ACP	MDSLNLMDFLVYLE[+448.143309].light	717.65484	3	369.18425	1	30

sp|P39135|SFP_BACSU	YSDLLAKDKDE.light	648.819532	2	705.34136	1	21.1
sp|P39135|SFP_BACSU	YSDLLAKDKDE.light	648.819532	2	634.30425	1	21.1
sp|P39135|SFP_BACSU	YSDLLAKDKDE.light	648.819532	2	506.20928	1	21.1

3.5.18. Targeted LC-MS/MS Data Availability

Raw data were imported into Skyline and are available on the LC-MS data sharing platform Panorama Public at the following link: https://panoramaweb.org/Structural%20control%20of%20bacterial%20acyl-CoA%20dehydrogenase.url.[278] Sample names for TcsD assays are labeled as "protein-substrate-replicate." For example, replicate one of the assay of wild type TcsD on pentanoyl-ACP is labeled "WT-val-01." Substrate abbreviations are as follows: val = pentanoyl-ACP, but = butyryl-ACP, 2pent = 2-pentenoyl-ACP, 2hex = 2-hexenoyl-ACP, 2hept = 2-heptenoyl-ACP, allylmal = allylmalonyl-ACP, propmal = propylmalonyl-ACP. The data for TcsD wild type assays on 2-pentenoyl-ACP, 2-hexenoyl-ACP, and 2-heptenoyl-ACP is located in the file named "WT assays_set 1_pub data.skyd." The data for TcsD wild type assays on propylmalonyl-ACP is located in the file named "Malonate assays_pub data.skyd." The data for TcsD wild type assays on butyryl-ACP and pentanoyl-ACP and the data for all mutant TcsD proteins on all substrates is located in the file named "Mutants assays_WT-but-val_pub data.skyd."
Sample names for *N. mexicana* ACAD assays are labeled as "substrate_replicate," and substrate abbreviations are the same as for TcsD assays. The data for butyryl-, pentanoyl-, and 2-pentenoyl-ACP are located in the file named "NmexACP_C4_C5_pub data.skyd," and data for 2-hexenoyl- and 2-heptenoyl-ACP are located in the file named "NmexACP_C6_C7_pub data.skyd."

3.5.19. Targeted LC-MS/MS Data Analysis: TcsD assays

After importing data, chromatograms were manually curated. Correct peaks were identified by comparison of their retention times to control peaks (i.e. the 2-pentenoyl-ACP active site peptide should elute before the pentanoyl-ACP active site peptide). For phosphopantetheine ejection peaks, only the transition derived from the peptide "DLGVDSLAMTELQAHALQR" was used for analysis and quantification. Peaks that were not correctly automatically selected were manually selected in Skyline and integrated using the built in Skyline integration function. For negative controls, an area corresponding to the same retention time as the peak of interest was integrated. After integrating, all phosphopantetheine ejection ion peak areas were normalized to an internal control peptide from the within the TcsA protein (a peptide that does not

contain catalytic residues) to account for differences in the amount of protein injected. The reported values in **Figure 3-3** were averages of three normalized replicate peak areas. Error bars represent standard deviations of each set of replicates.

3.5.20. Targeted LC-MS/MS Data Analysis: *N. mexicana* ACAD assays

After importing data into Skyline, chromatograms were manually curated (including manual integration where necessary) as described above for TcsD assays. Plots in Figures **3-5E** and **3-5F** and **Figure 4-13** correspond to the LC-MS/MS chromatograms of the phosphopantetheine ejection transition from the parent peptide "MDSLNLMDFLVYE" in the +2 charge state.

3.5.21. Preparation of denatured TcsD supernatants for untargeted LC-MS analysis

TcsD was expressed and purified via nickel affinity chromatography as described above. The highest purity nickel affinity elution fractions were pooled, and protein was concentrated to approximately 17 mg/mL or 375 μM (determined by absorbance using a Nanodrop and the molar absorbance of TcsD). After concentration, 150 μL of protein was denatured by either boiling for 10 minutes or the addition of 300 μL of LC grade acetonitrile. The denatured protein was pelleted by centrifuging at maximum speed in a benchtop centrifuge for 2 minutes. Supernatants were removed and lyophilized overnight. After lyophilization, supernatants were resuspended in 75 μL of LC grade water for an effective metabolite concentration of 750 μM. Samples were diluted to an effective metabolite concentration of 50 μM in 50:50 (v/v) LC grade water:methanol prior to untargeted LC-MS analysis. Standards of flavin adenine dinucleotide (FAD), Coenzyme A, pantetheine, pantothenate, and butyryl-CoA were prepared in 50:50 (v/v) LC grade methanol:water to a concentration of 20 μM.

3.5.22. High resolution untargeted LC-MS analysis of denatured TcsD supernatants

Supernatants of denatured TcsD samples were analyzed using an Agilent 6545 LC-QTOF system as previously described.[279] Analytes were separated using a SeQuant ZIC hydrophilic interaction chromatography (HILIC) column (150 mm length, 4.6 mm internal diameter, 5 μm particle size) coupled to a SeQuant ZIC-pHILIC guard column (20 mm length, 2.1 mm internal diameter, 5 μm particle size). The mobile phase consisted of 10 mM ammonium carbonate and 118.4 mM ammonium hydroxide in acetonitrile-water (60.2:39.8 v/v). Analytes were separated isocratically using the following pump parameters:

Table 3-10 LC gradient for untargeted LC-MS of denatured TcsD supernatants

Time (min)	Mobile phase (%)	Flow rate (mL/min)
0	100	0.45

6	100	0.45
6.5	100	0.605
12.5	100	0.605

The mass spectrometer source was operated with the following parameters: Gas Temperature = 300°C, Drying Gas = 10 L/min, Nebulizer Pressure = 20 psi, Sheath Gas Temperature = 350°C, Sheath Gas Flow = 12 L/min, and Capillary Voltage = 3500 V, Nozzle Voltage = 2000 V. The TOF mass spectrometer was programmed to operated in scan mode with the following parameters: Fragmentor Voltage = 100 V, Skimmer Voltage = 50 V, Oct 1 RF Vpp = 400 V, Mass Range = 70 m/z – 1100 m/z, Acquisition Rate = 0.86 spectra/s, Acquisition Time = 1162.8 ms/spectrum, and Transients/spectrum = 9532 (negative ion mode) or 9485 (positive ion mode). All samples were analyzed in both negative ion mode and subsequently in positive ion mode. Data analysis was performed in Agilent MassHunter Qualitative Analysis (v B.05.00).

3.5.23. Substrate modeling into TcsD active site

Substrates for modeling were drawn using ChemDraw, and ligand restraints were generated using eLBOW in the Phenix software suite. Substrates were manually placed into the TcsD active site using Coot. Real Space Refinements of the substrates were performed using the unknown density in the substrate-binding region of the enzyme active site. The refinements resulted in the generation of both *cis* and *trans* isomers of the substrate with planar fatty acyl tails. After substrate generation, the TcsD structure was aligned with the structures of ACADs which had been co-crystallized with thioester substrates. Alignments were performed using the least squares fit LSQ superpose function of Coot. Proteins were aligned using the 3-residue loop containing the catalytic glutamate and the residues immediately upstream and downstream of the glutamate (**Table 3-11**). These alignments were used for analysis of the relative positioning of FAD cofactors and for substrate modeling. After alignment, the 2-pentenoyl substrates were manually translated and rotated into a position within the TcsD active site that corresponds to approximately the equivalent position of substrates in the aligned homolog structures. The feasibility of hydrogen bonding interactions was confirmed using the Measure tool in Coot considering a maximum distance of 3.5Å between heteroatoms. Substrate lengths were also determined using the Measure tool to calculate the distance from the carbonyl carbon of a given molecule to the distal carbon of the fatty acyl tail. The length of the 2-hexenoyl substrate was calculated using a substrate molecule that was generated using Ligand Builder in Coot. The approximate length of the 2-heptenoyl substrate was calculated by adding half the distance between

the fourth and sixth (saturated) carbons of the 2-hexenoyl substrate to the total length of the 2-hexenoyl substrate.

Table 3-11 ACAD structures used in alignment for substrate modeling

PDB code	Host organism	Protein description	Amino acids used for LSQ alignment
6U1V	*Streptomyces tsukubensis* NRRL 18488	TcsD	Ile 363
			Glu 364
			Gly 365
1jqi	*Rattus norvegicus*	Rat short chain ACAD	Tyr 367
			Glu 368
			Gly 369
2vig	*Homo sapiens*	Human short chain ACAD	Tyr 391
			Glu 392
			Gly 393
1udy	*Sus scrofa*	Pig medium chain ACAD	Tyr 375
			Glu 376
			Gly 377
1buc	*Megasphaera elsdenii*	Bacterial butyryl-CoA ACAD	Tyr 366
			Glu 367
			Gly 368
3mdd	*Sus scrofa*	Pig medium chain ACAD	Tyr 375
			Glu 376
			Gly 377

3.5.24. Bioinformatic identification of TcsD homologs

First, a Hidden Markov Model (HMM) that describes γ,δ-ACADs was generated using the sequences of several TcsD enzymes from various FK506 producers and HliR (**Table 3-12**).[264] The HMM was then used to search for homologs among all proteins in the Uniprot database, after which the genomes of organisms that contain potential γ,δ-ACADs were analyzed using antiSMASH.[280] An additional search was performed by querying the results of the HMM search against GenBank using protein BLAST.[263] Finally, CORASON-BGC was used for the identification of the syntenic genomic contexts of the identified genes.[262] The results from each of these searches were then manually curated based on the presence of key motifs identified through the structural and biochemical characterization of TcsD.

Table 3-12 Protein sequences used to build γ,δ-ACAD HMM used in genome mining

GenBank Accession Number	Host organism	Sequence
ADU56309.1	*Streptomyces sp*. KCTC 11604BP	MSESERLGIVRDFVAREILGREGILDSLADAPLALYERFAETGL MNWWVPKEHGGLGLGLEESVRIVSELAYGDAGVAFTLFLPV LTTSMIGWYGSEELKERFLGPLVARRGFCATLGSEHEAGSELA RISTTVRRDGDTLVLDGTKAFSTSTDFARFLVVIARSADDPAR YTAVTVPRDAPGLRVDKRWDVIGMRASATYQVSFSDCRVP GDNALNGNGLRLLEIGLNASRILIAASALGVARRIRDVCMEY GKTKSLKGAPLVKDGVFAGRLGQFEMQIDVMANQCLAAAAR AYDATAARPDAARVLLRQGAQKSALTAKMFCGQTAWQIAS TASEMFGGIGYTHDMVIGKLLRDVRHASIIEGGDDVLRDLVY QRFVVPTAKRT
WP_055551085.1	*Streptomyces kanamyceticus*	MSEPEHLDTVRKFVAQEVLGRETHLDSLADAPLALYERFAET GLMNWWVPEEHGGLGLGLEDSVRIVSELAYGDAGVAFTLFL PILTTSMVSWYGSAELKEKLLDPLVAHRGFCATLGSEHEAGSE LAKISTVVRRDGEGLVLDGTKAFSTSTDFAQFLVVIARSAENP TRYLAVAVERDAPGLRIDKRWDVIGLRASATYQVSFSDCHVP AGNALDGHGLRLLEIGLNASRILIAATALGVARRIRDLCMEYA KTKSLKGAPLVNDAVFAGRLGQFEMQIEVMANQCLAAART YDATAARPDAARTLLRQGAQKSALTAKMFCGQTAWQIAST ASEMFGGIGYTHDVPIGKLLRDVRHASIIEGGDDVLRDLVFH RFVVPTAKRT

WP_006350839.1	*Streptomyces tsukubensis*	MSESERLGIVRDFVAREILGREGILDSLADAPLALYERFAETGLMNWWVPKEHGGLGLGLEESVRIVSELAYGDAGVAFTLFLPVLTTSMIGWYGSEELKERFLGPLVARRGFCATLGSEHEAGSELARISTTVRRDGDTLVLDGTKAFSTSTDFARFLVVIARSADDPARYTAVTVPRDAPGLRVDKRWDVIGMRASATYQVSFSDCRVPGDNALNGNGLRLLEIGLNASRILIAASALGVARRIRDVCMEYGKTKSLKGAPLVKDGVFAGRLGQFEMQIDVMANQCLAAAARAYDATAARPDAARVLLRQGAQKSALTAKMFCGQTAWQIASTASEMFGGIGYTHDMVIGKLLRDVRHASIIEGGDDVLRDLVYQRFVVPTAKRT
AMM72019.1	*Haliangium ochraceum* DSM 14365	MSADTTKKNPLIEPIRGFVREHVLGREQQLDAGGELPLDIYEAFRKAGLANWWLPESYGGHGLSLEQSVDIVAELAYGDAGLAFAFFLPILSTSVIEQFGSEEQKQRYLPALAKSGGSCATMASEEKAGSELIRTEALARGSAEEGFKLSGDKYFSTNADTAELLIVYARIAGPTPAYGAFLVPRSADGIRIVRRWEMNGLRASGTYELELRDCPAESQLAGNGLRILEVGLNSSRTLMAACAVGIARRVRDVCLGYARNKEIKNTKLFNNHVFGAKLGQMEAELDGMMAVCRTAAREMDEIASREDAAKVFFREGTLKSVIVAKMLCGQLGWKIASVGSESLGGLGYTHDSIIGKLLRDVRYVSLVEAGDDVLRDLVFSRHVLPRFMSEIE

3.5.25. Sequence alignment of TcsD homologs

A sequence alignment of TcsD, HliR, and several α,β-ACADs was generated using Clustal Omega.[281] The sequence alignment was visualized with respect to the TcsD structure using ESPript 3.0.[282] To calculate the positional occupancy of amino acids within the γ,δ-ACADs, a sequence alignment of the identified proteins was generated using MAFFT.[283] Positional occupancies were calculated using the bioinformatics tool UGene.[284]

3.5.26. Position weight matrix

The Shannon entropy of the aligned γ,δ-ACADs was calculated at each position within a sequence alignment. All of the identified γ,δ-ACAD sequences were aligned using MAFFT.[283] The MSA was then refined so that all positions maintained at least 10% occupancy using the ProDy python library. The Shannon entropy was then calculated using ProDy.[285,286] The sequence logo was generated using Weblogo (https://weblogo.berkeley.edu/logo.cgi).

3.6. Miscellaneous

3.6.1. Acknowledgements

This work was performed as part of the DOE Joint BioEnergy Institute (https://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). In addition, this work was supported by the Agile BioFoundry (https://agilebiofoundry.org), through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. J.M.B. was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1106400.

3.6.2. Competing interests

JDK has a financial interest in Amyris, Lygos, Demetrix, Constructive Biology, Maple Bio, and Napigen.

3.6.3. Structural data deposition

The atomic coordinates and structural factors of TcsD have been deposited in the Protein Data Bank, PDB ID code 6U1V.

3.6.4. LC-MS/MS data deposition

All targeted LC-MS/MS data from TcsD assays has been uploaded to Panorama Public[278] and is publicly available at the following link: https://panoramaweb.org/Structural%20control%20of%20bacterial%20acyl-CoA%20dehydrogenase.url

Chapter 4. Supplementary information for Chapter 3

4.1. LC-MS/MS Phosphopantetheine ejection assay method development

In order to assay TcsD activity on TcsA-bound substrates, we first developed a targeted proteomics assay for the identification of substrates tethered to the phosphopantetheinyl arm of the acyl carrier proteins (ACP) domain of TcsA. In a typical "phosphopantetheine ejection" experiment,[255] a protein of interest is digested with trypsin and the digested peptides are analyzed via LC-MS/MS (main text Figure 2A). However, after digesting TcsA with trypsin we were unable to detect an ion representing the ACP active site peptide, most likely because the large size (~40 residues) of the tryptic peptide placed it outside of the detectable range of our mass spectrometer (**Figure 4-1**). Similarly, the active site peptide could not be detected after digestion with another commonly-used protease, chymotrypsin. In order to obtain an appropriately-sized peptide, we performed a sequential digestion of TcsA with two proteases, Asp-N then trypsin, and identified the resulting "holo" and "acyl" active site peptides using high resolution untargeted LC-MS (**Figure 4-1** & **4-2**). Surprisingly, Asp-N did not consistently cleave at the predicted residue D373 of the conserved "DSL" motif of the acyl carrier protein (ACP) of TcsA but instead at the preceding D369, possibly due to steric effects arising from the presence of the phosphopantetheine arm appended to S374.

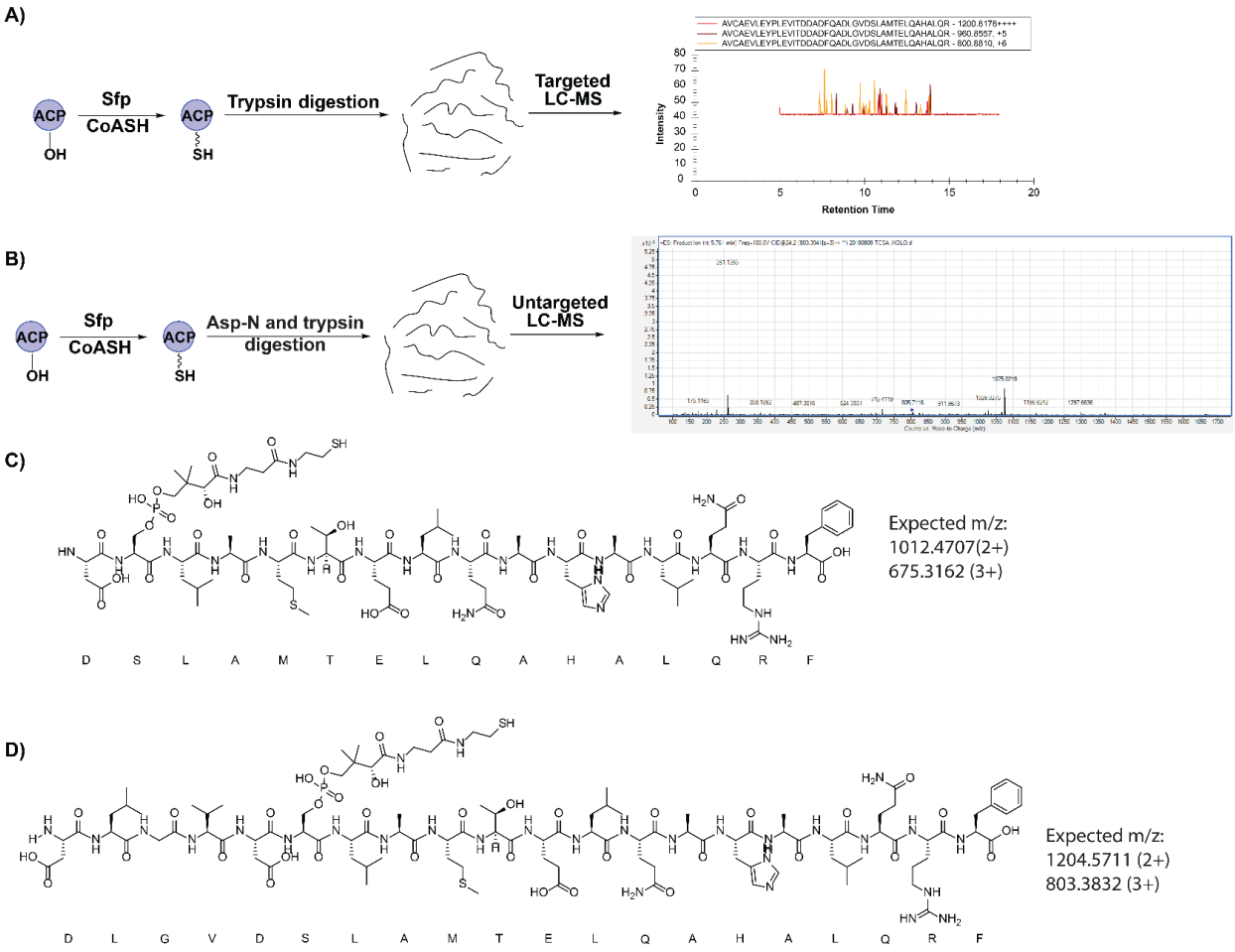


Figure 4-1 Development of a phosphopantetheine ejection method for studying intermediates bound to TcsA-ACP. A) Digestion of TcsA with trypsin results in an active site peptide 40 amino acids in length which is not detectable using targeted LC-MS/MS. **B)** Dual digestion of TcsA with Asp-N and trypsin results in a detectable peptide from which phosphopantetheine ejection can be observed. **C)** Predicted TcsA active site peptide resulting from TcsA digestion with Asp-N and trypsin. **D)** Observed TcsA active site peptide resulting from TcsA digestion with Asp-N and trypsin (verified with data in Figure 4-2).

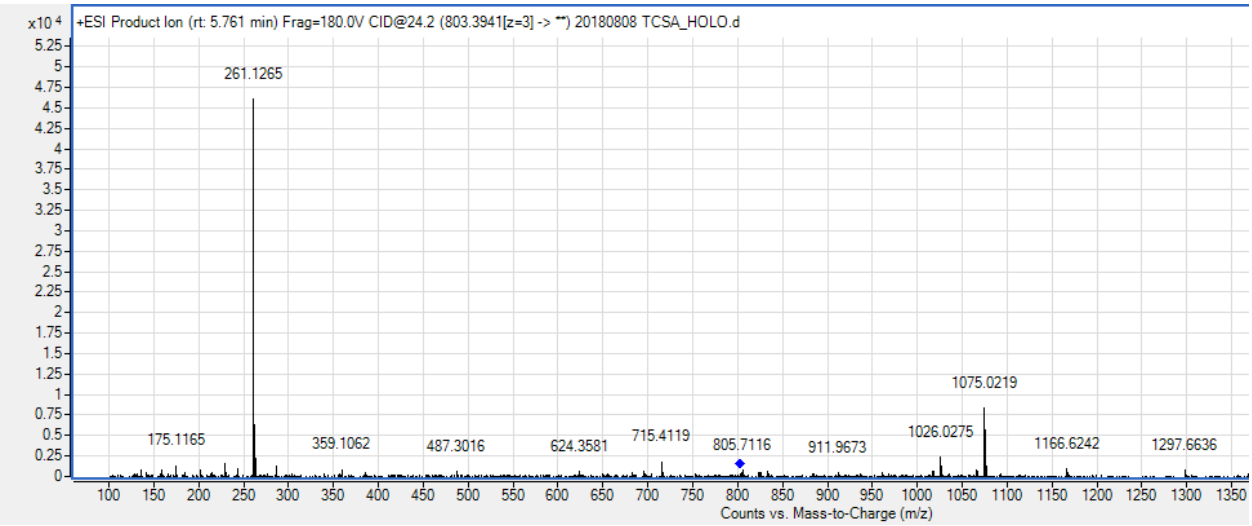


Figure 4-2 Product ion mass spectrum showing phosphopantetheine ejection ion resulting from holo-TcsA digested with Asp-N and trypsin. The product ion is derived from a parent ion with m/z = 803.394 [z=3], representing the peptide "DLGVDSLAMTELQAHALQR" in which Asp-N misses one cleavage, not the expected "DSLAMTELQAHALQR" active site peptide (see Figure 4-1).

4.2. Unidentified density in active site of TcsD crystal structure

An unknown metabolite consistently co-crystallized in the TcsD active site, occupying the space where substrates are predicted to bind (**Figure 4-3**). We attempted to identify this unknown density by performing untargeted high resolution LC-MS of denatured protein samples (**Figures 4-4** & **4-5**), but were unable to identify any unique masses in the supernatants of denatured TcsD samples. We also compared the denatured protein supernatant LC-MS chromatograms to several standards, including flavin adenine dinucleotide (FAD), Coenzyme A, pantetheine, pantothenate, and butyryl-CoA, but only the m/z corresponding to FAD could be detected in extracted ion chromatograms of either sample. We additionally used the Ligand Identification function of the Phenix software suite,[287,288] attempting to model 180 of the most common ligands from the Protein Data Bank into the density, but none of the ligands fit appropriately.

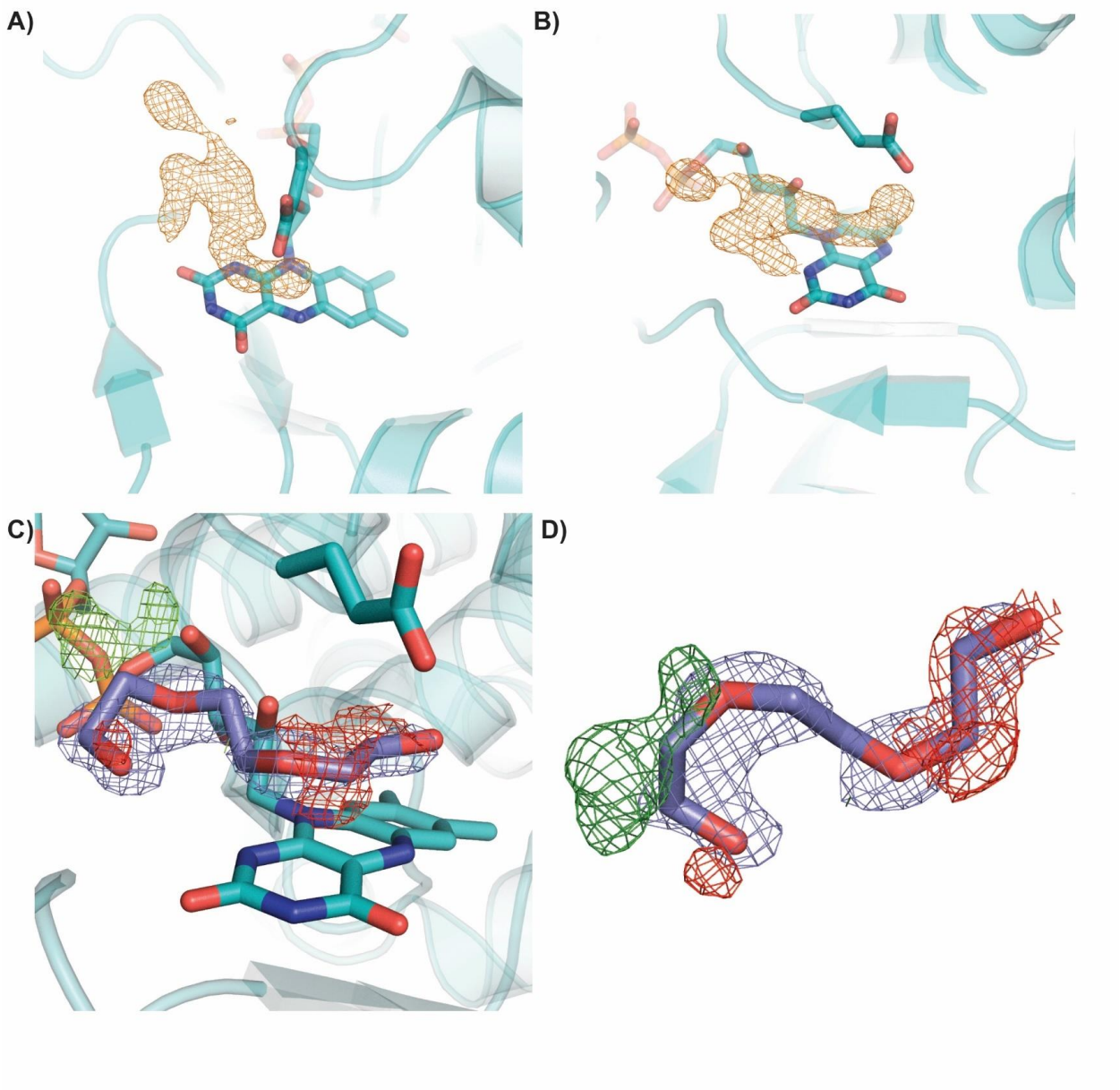


Figure 4-3 Unidentified density in active site of TcsD and modeling of PEG trimer into density. A) and B) Shape of the unidentified density (in orange mesh) in the TcsD active site ($F_o - F_c$ map at 3σ) **C) and D)** Model of TcsD after refinement with a PEG trimer (purple sticks) occupying the unknown density, showing that this PEG fragment does not fit properly within the density. The $2F_o - F_c$ map is shown in blue mesh at 1σ. The $F_o - F_c$ map is shown in green at 3σ and in red at -3σ.

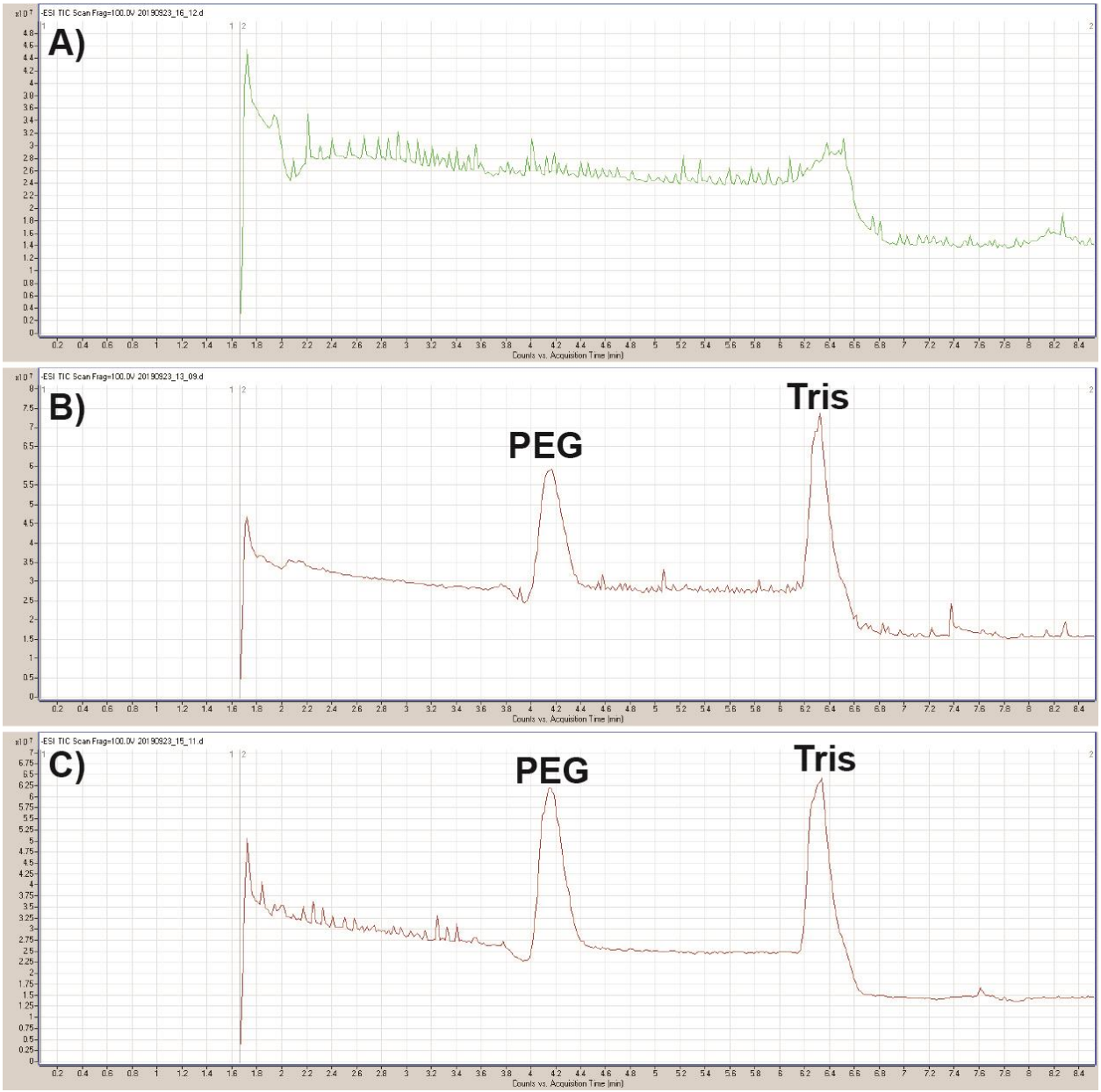


Figure 4-4 High resolution untargeted LC-TOF analysis of denatured purified TcsD samples in negative ion mode. Samples are as follows: **A)** Crystallization buffer **B)** Supernatant from boiled TcsD **C)** Supernatant from acetonitrile-denatured TcsD. Peaks pertaining to PEG (a common LCMS contaminant) and Tris are labeled.

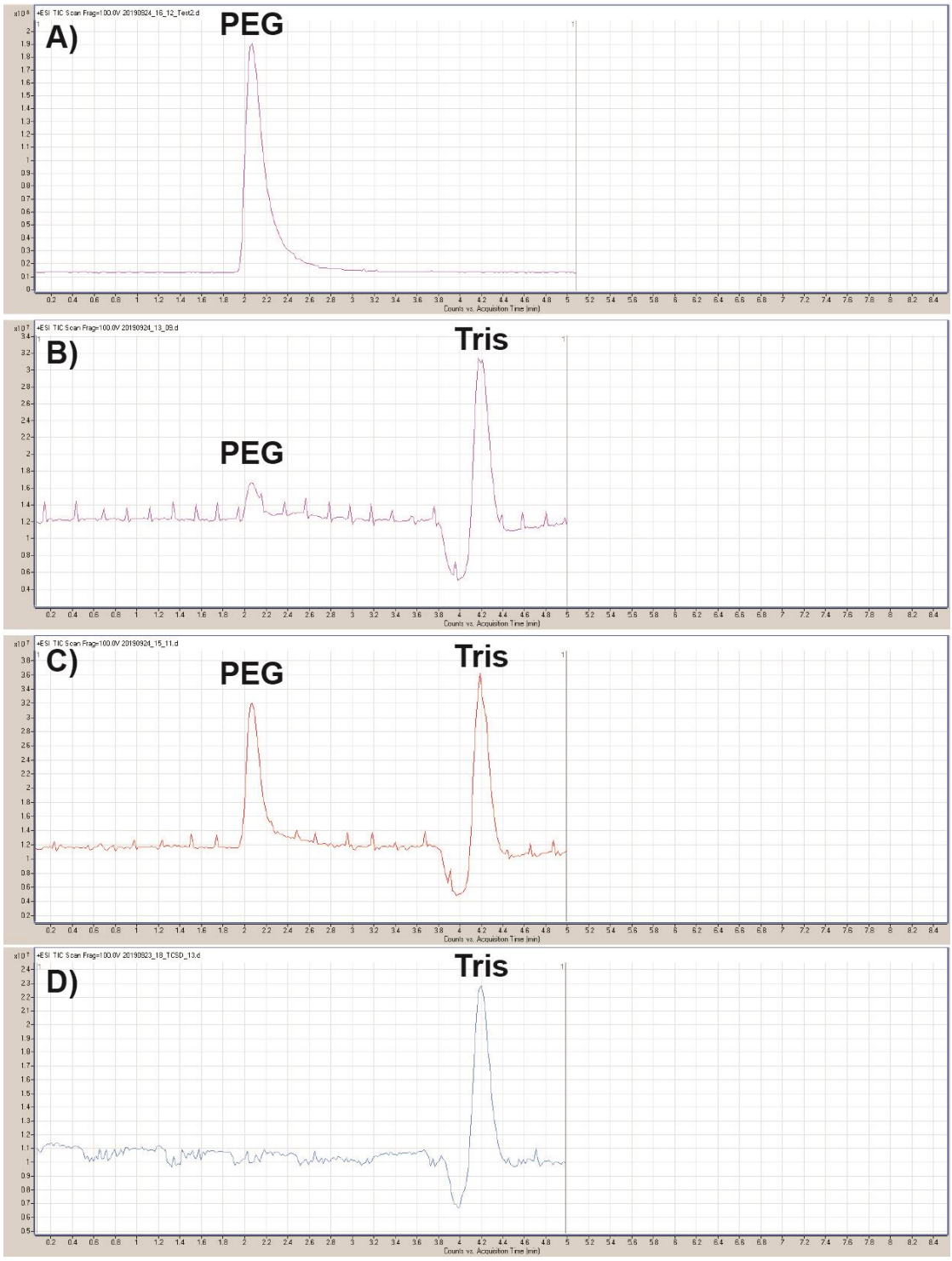


Figure 4-5 High resolution untargeted LC-TOF analysis of denatured purified TcsD samples in positive ion mode. Samples are as follows: **A)** Crystallization buffer **B)** Supernatant from boiled TcsD **C)** Supernatant from acetonitrile-denatured TcsD, **D)** lysis buffer. Peaks pertaining to PEG (a common LCMS contaminant) and Tris are labeled.

4.3. Supplementary Figures and Tables

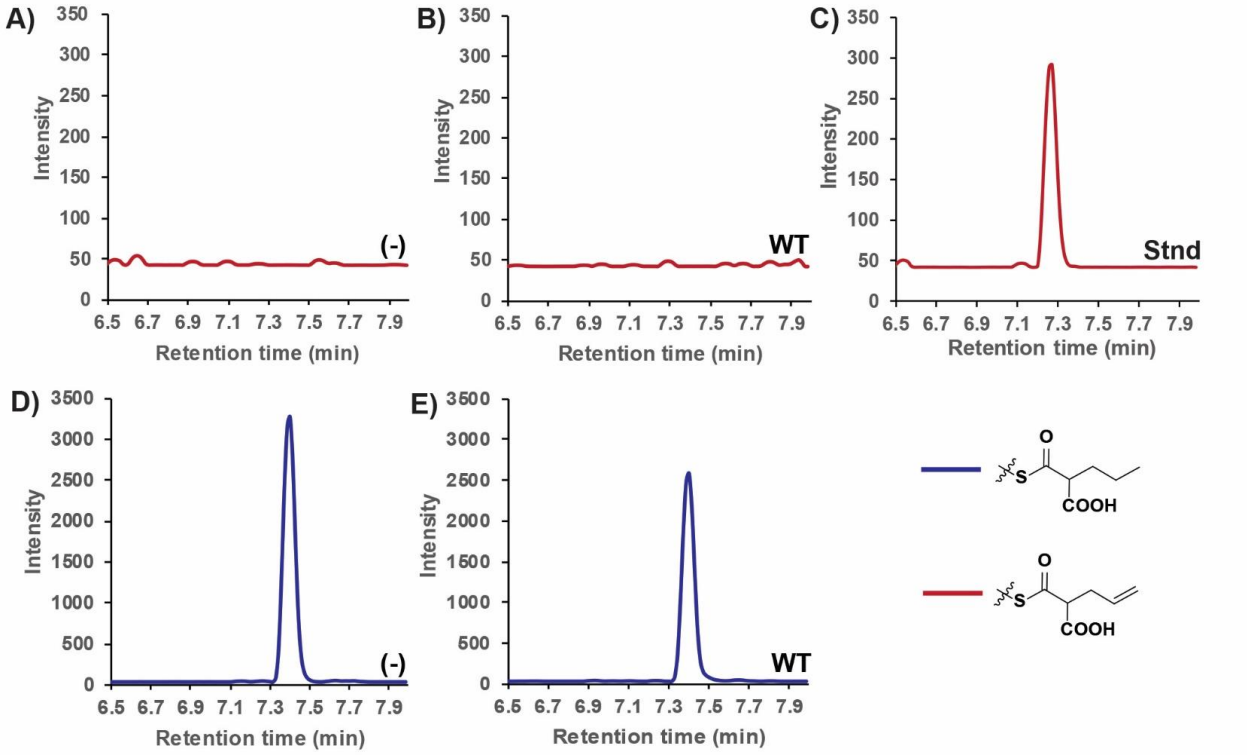


Figure 4-6 TcsD activity on propylmalonyl-ACP (pathway *A2* from Figure 1) analyzed via targeted LC-MS/MS. WT = wild type TcsD, (-) = negative control, Stnd = standard. The chromatograms shown depict: **A)** the allylmalonyl-ACP transition for the negative control reaction **B)** the allylmalonyl-ACP transition for the wild type TcsD reaction **C) a**n allylmalonyl-ACP standard **D)** the propylmalonyl-ACP transition for the negative control reaction **E)** the propylmalonyl-ACP transition for the TcsD wild type reaction

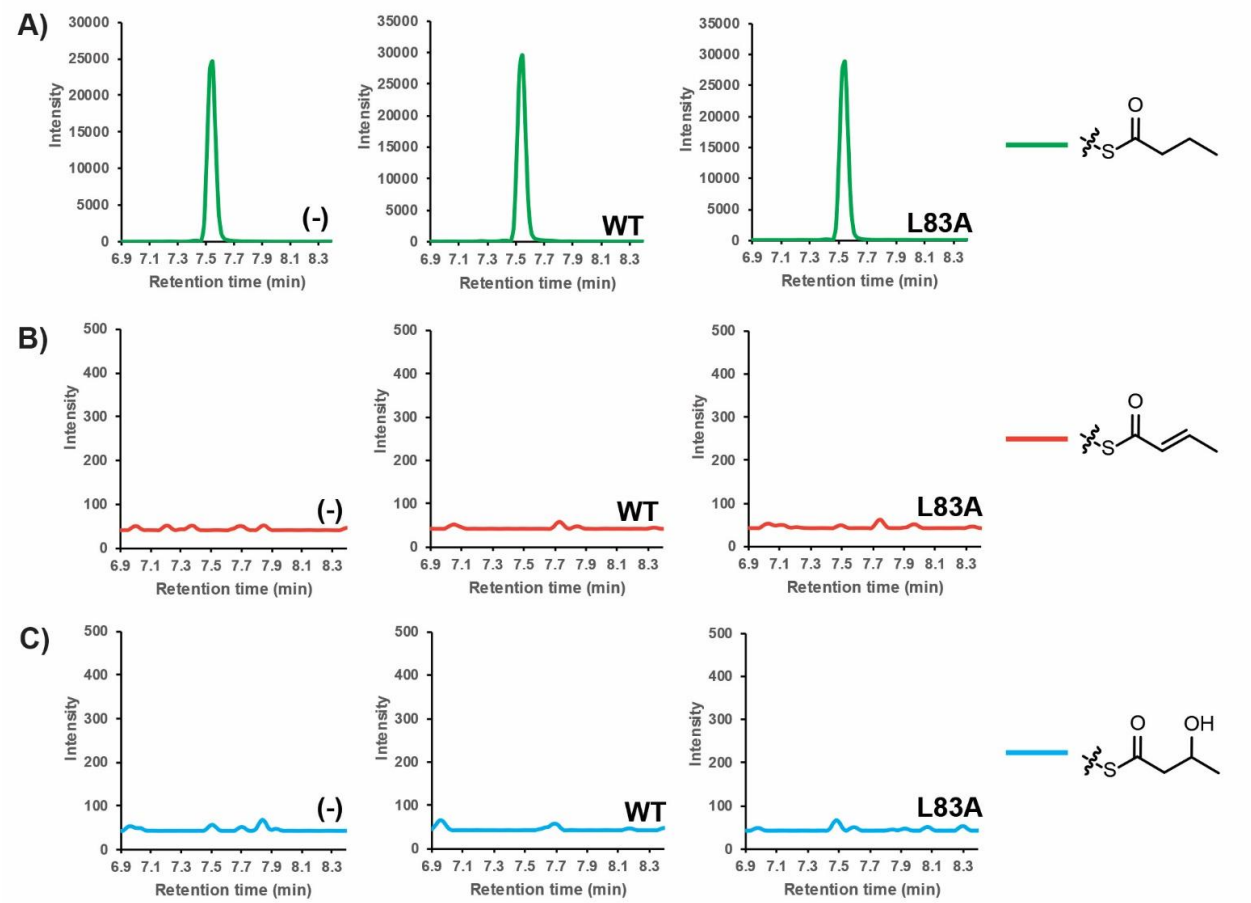


Figure 4-7 Activity of TcsD wild type and L83A mutant on butyryl-ACP analyzed by targeted LC-MS/MS. WT = wild type TcsD, (-) = negative control, L83A = TcsD L83A mutant. Chromatograms representing different transitions are color coded (key is shown on right side of figure). The chromatograms shown depict: **A)** butyryl-ACP, **B)** crotonyl-ACP, or **C)** 3-hydroxybutyryl-ACP for each assay (negative control, TcsD wild type, or TcsD L83A).

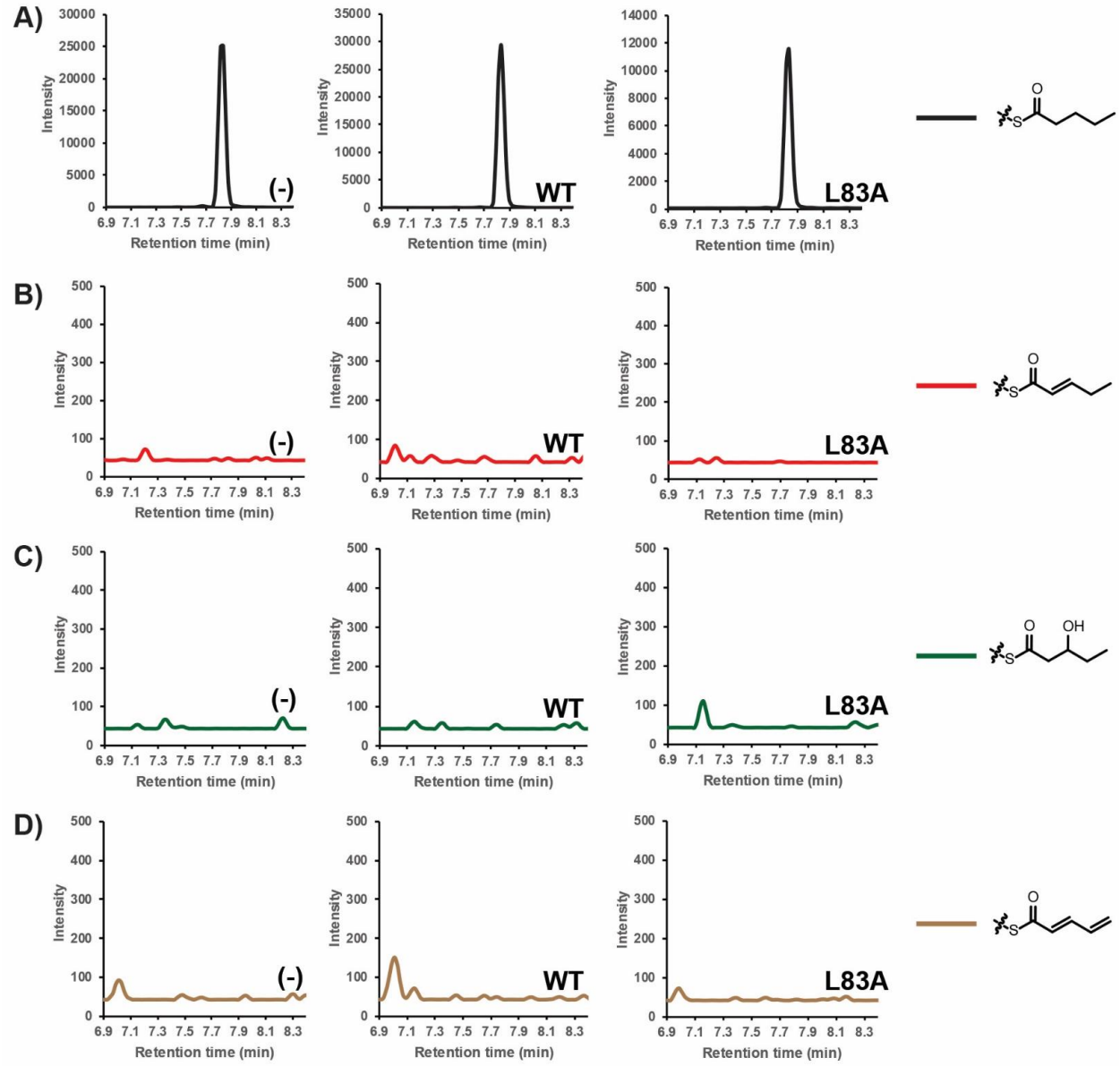


Figure 4-8 Activity of TcsD wild type and L83A mutant on pentanoyl-ACP analyzed by targeted LC-MS/MS. WT = wild type TcsD, (-) = negative control, L83A = TcsD L83A mutant. Chromatograms representing different transitions are color coded (key is shown on right side of figure). The chromatograms shown depict: **A)** pentanoyl-ACP, **B)** 2-pentenoyl-ACP, **C)** the 3-hydroxypentanoyl-ACP, or **D)** 2,4-pentadienoyl-ACP for each assay (negative control, TcsD wild type, or TcsD L83A).

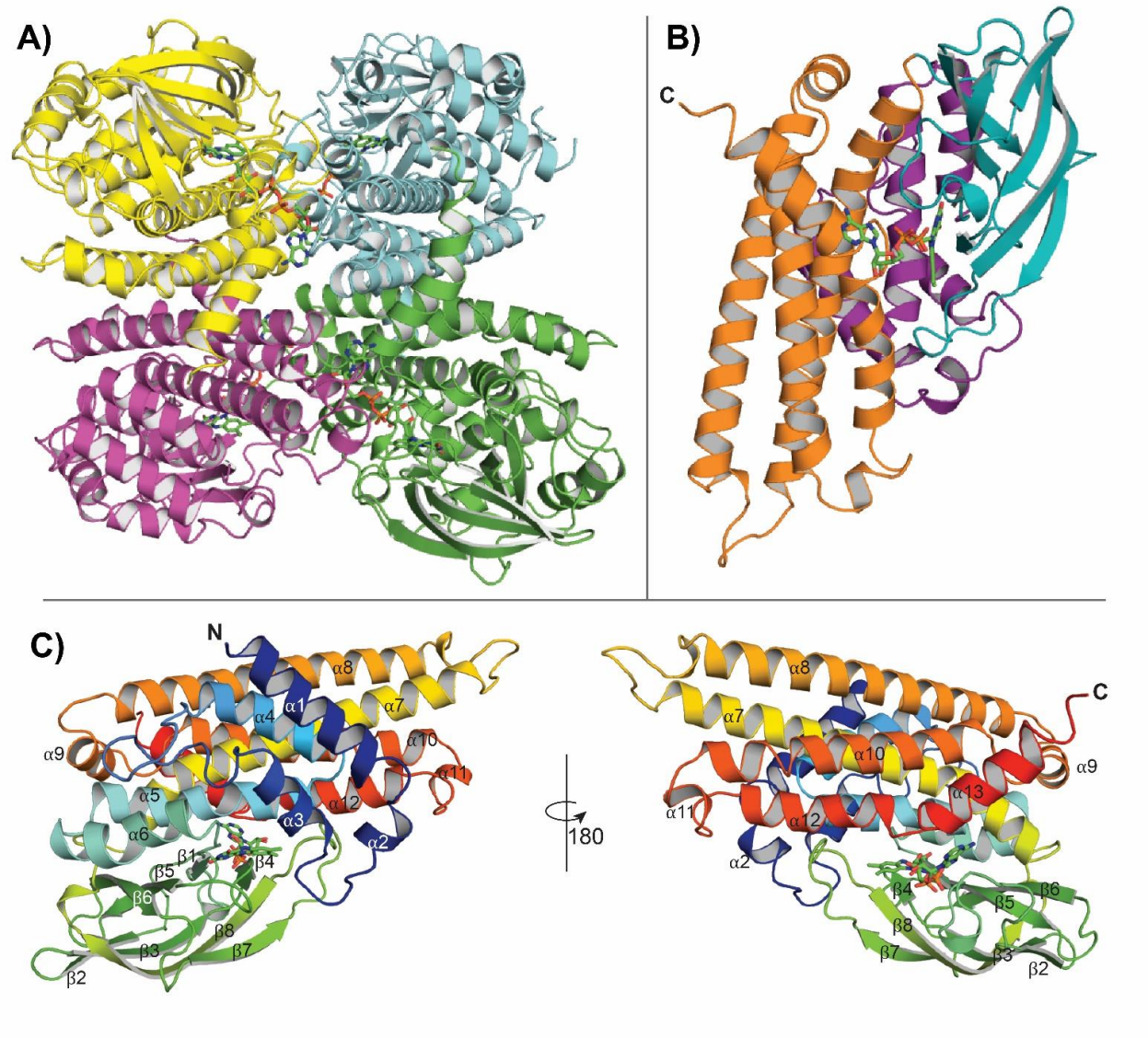


Figure 4-9 Overall structure of TcsD. A) Tetrameric structure of TcsD. Each subunit is shown in a different color. FAD cofactors are shown as sticks. **B)** A single TcsD subunit. The N-terminal α-helix domain (purple), middle β-sheet domain (teal), and C-terminal α-helix domain (orange) are highlighted. C) A single TcsD monomer colored in a progressive rainbow from its N- (blue) to C-terminus (red). Helices and sheets are numbered.

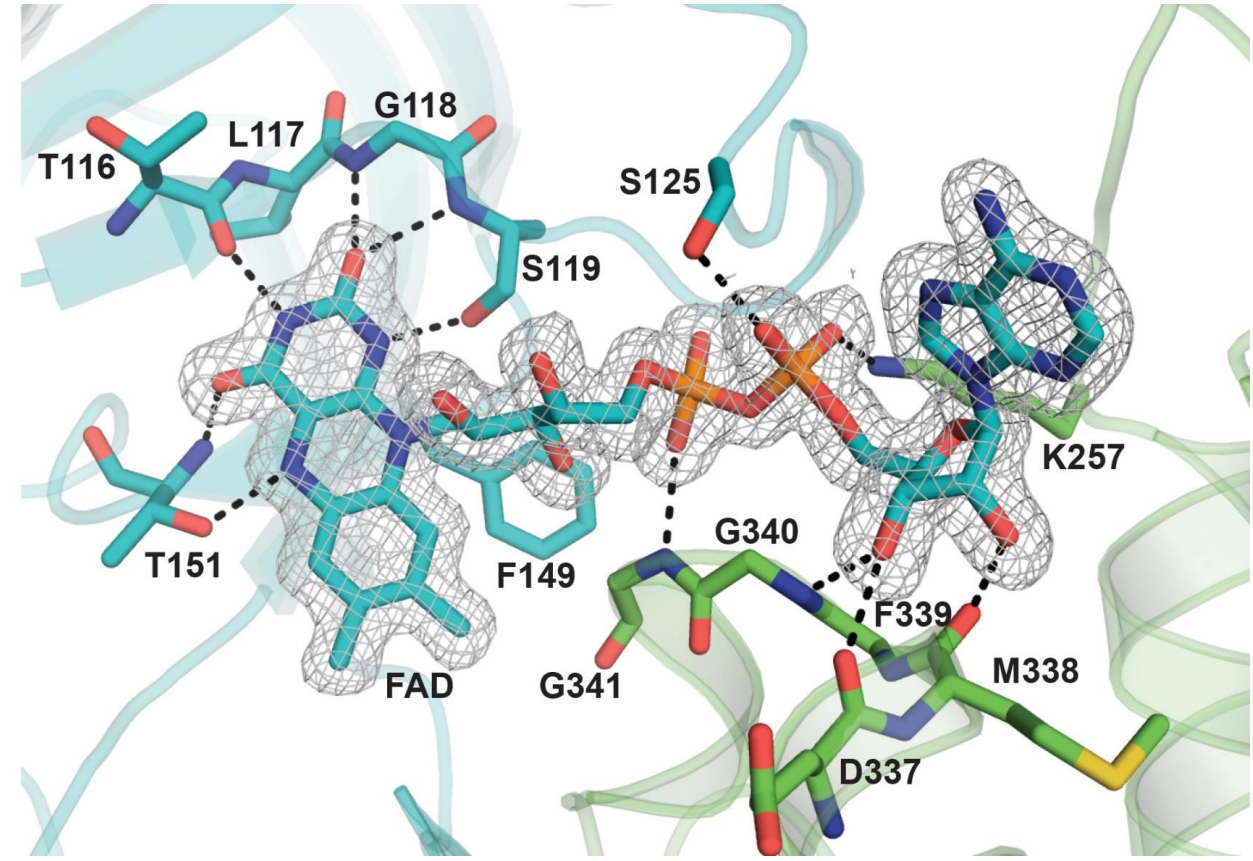


Figure 4-10 FAD-interacting residues of TcsD. The $2F_o - F_c$ electron density map of FAD is shown with gray mesh and is contoured at 1σ. Residues that interact with FAD via hydrogen bonds are shown as sticks, with hydrogen bonds depicted as black dashed lines. Residues pertaining to the same subunit as FAD (T116, L117, G118, S119, S125, F149, T151) are shown in blue, while residues pertaining to the adjacent subunit (K257, D337, M338, F339, G340, G341) are shown in green.

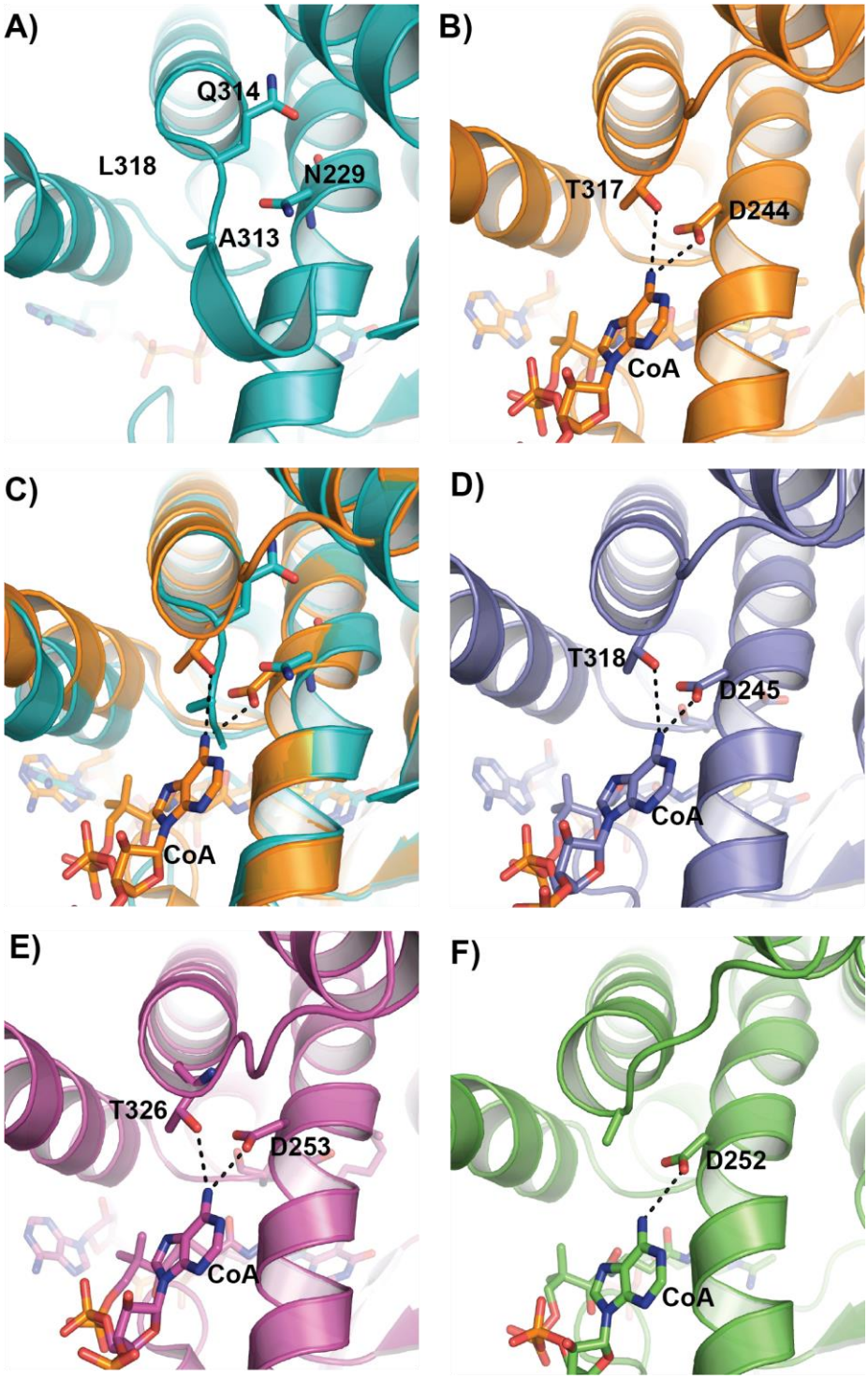


Figure 4-11 Residues of ACADs that hydrogen bond with the nucleotide portion of CoA. Showing corresponding region in **A)** TcsD, **B)** *M. esldenii* butyryl-CoA dehydrogenase (1buc), **C)** *M. esldenii* butyryl-CoA dehydrogenase overlaid with TcsD, **D)** *Sus scrofa* medium chain ACAD (1udy), **E)** rat short chain ACAD (1jqi), **F)** human isovaleroyl-CoA dehydrogenase (1ivh). TcsD has an Asn residue in the place of the

conserved Asp of α,β-ACADs. Helix 9 occupies some of the space where the nucleotide portion of CoA would normally bind. The conserved Thr residue is replaced with a Leu residue.

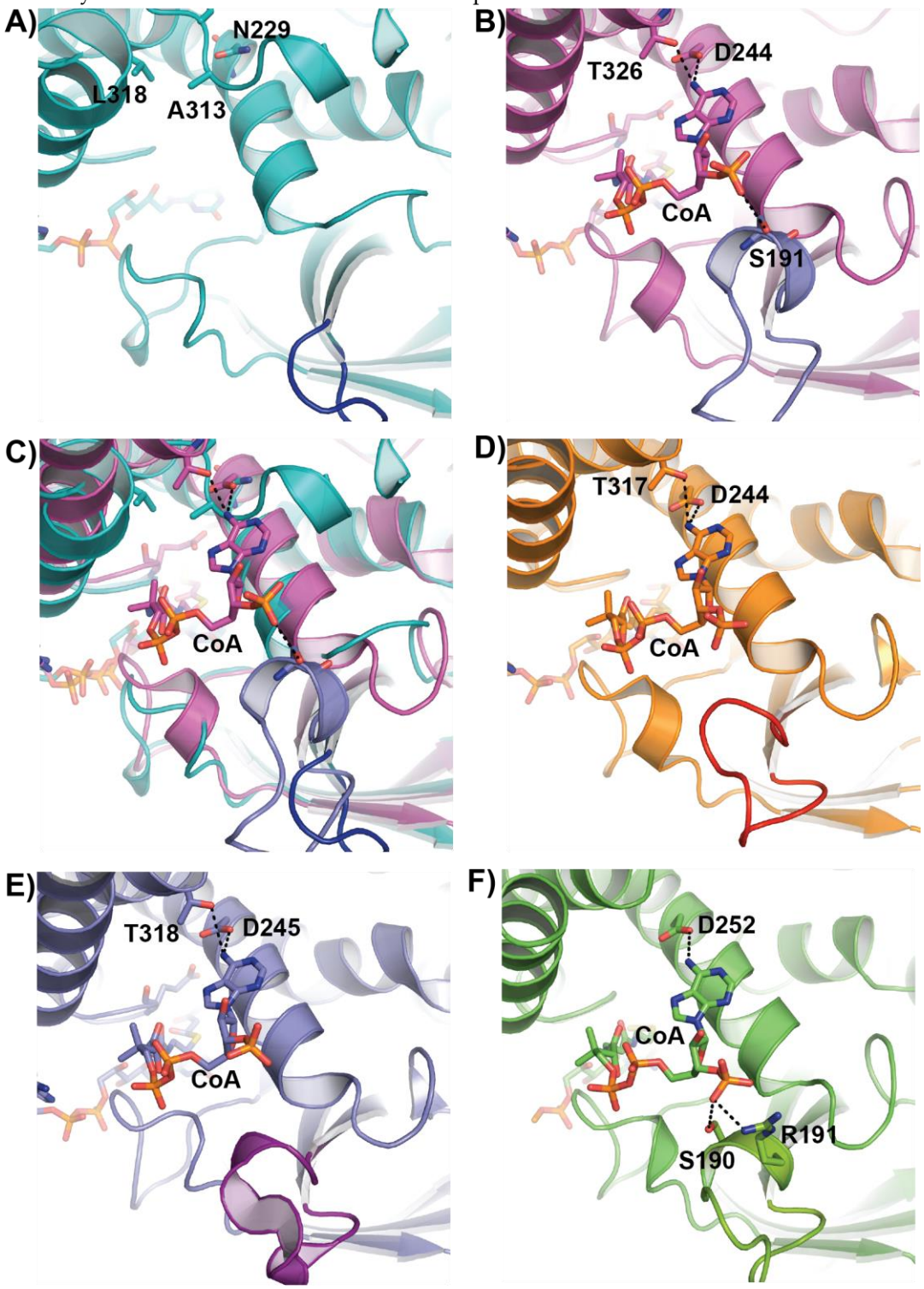


Figure 4-12 CoA phosphate-binding region of acyl-CoA dehydrogenases and TcsD. Showing corresponding regions with the structures of: **A)** TcsD, **B)** rat short chain ACAD (1jqi), **C)** rat short-chain

ACAD overlaid with TcsD, **D)** *M. esldenii* butyryl-CoA dehydrogenase (1buc), **E)** *Sus scrofa* medium chain ACAD (1udy), **F)** human isovaleroyl-CoA dehydrogenase (1ivh). The loop that approaches the phosphate groups is colored in a different shade in each panel.

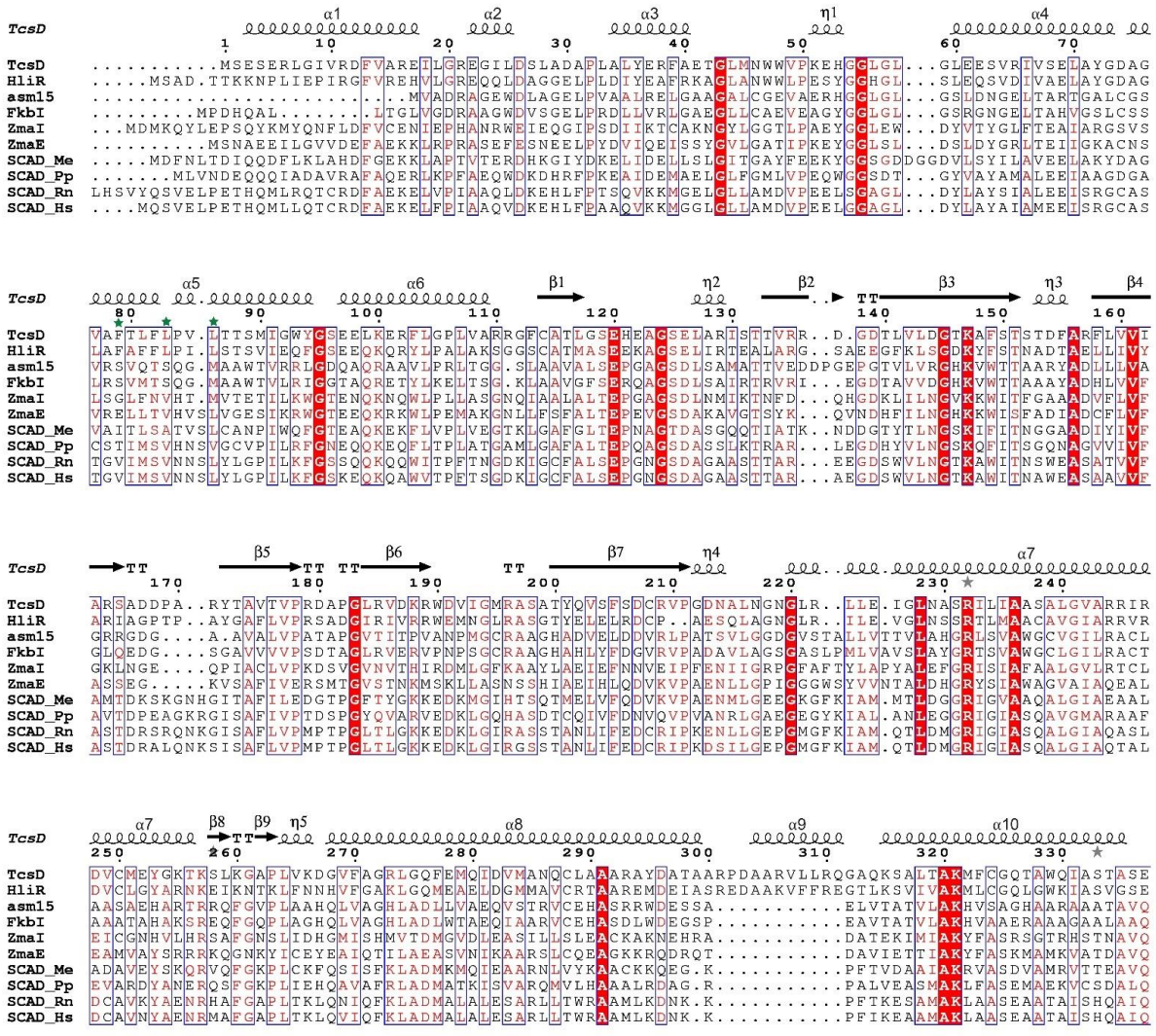


Figure 4-13 Sequence alignment of TcsD with homologs. Structural elements noted above the alignment are based on the TcsD structure. Proteins included in the alignment and their Genbank accession numbers are: γ,δ-ACADs TcsD (ADU56309.1) and HliR (AMM72019.1), hydroxymalonate semialdehyde

dehydrogenases asm15 (AAM54093.1), FkbI (TAI41666.1), and ZmaE (AAD40109.1), aminomalonate semialdehyde dehydrogenase ZmaI (AAR87758.1), and α,β-ACADs SCAD_Me (pdb 1buc), SCAD_Pp (NP_744365.1), SCAD_Rn (pdb 1jqi), and SCAD_Hs (pdb 2vig). Green stars denote residues that form the active site pocket.

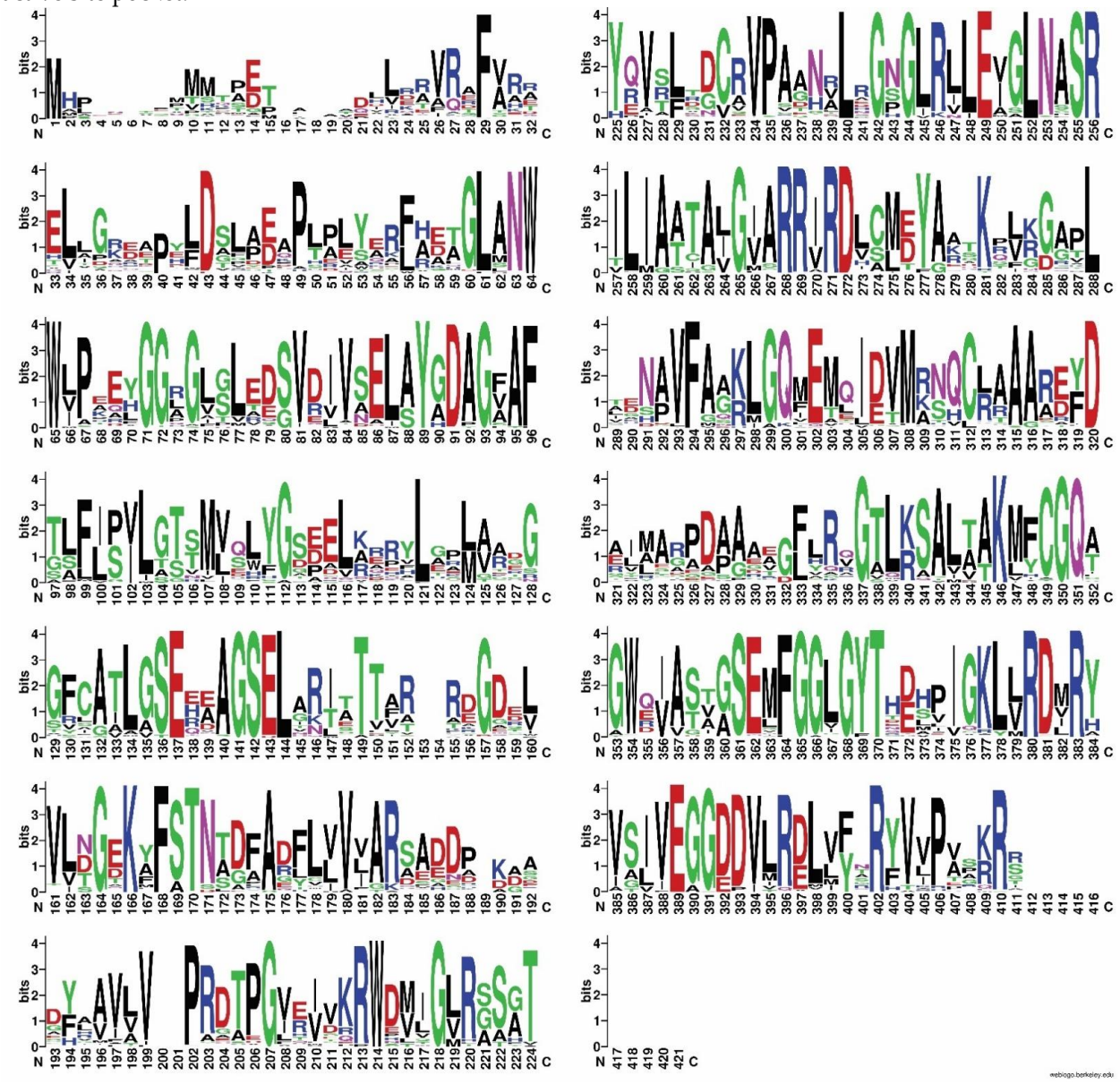


Figure 4-14 Web logo depicting the conservation of amino acids across a multiple sequence alignment of all γ,δ-ACADs identified in this work. Larger-sized letters represent higher conservation, while smaller letters represent less conserved amino acids.

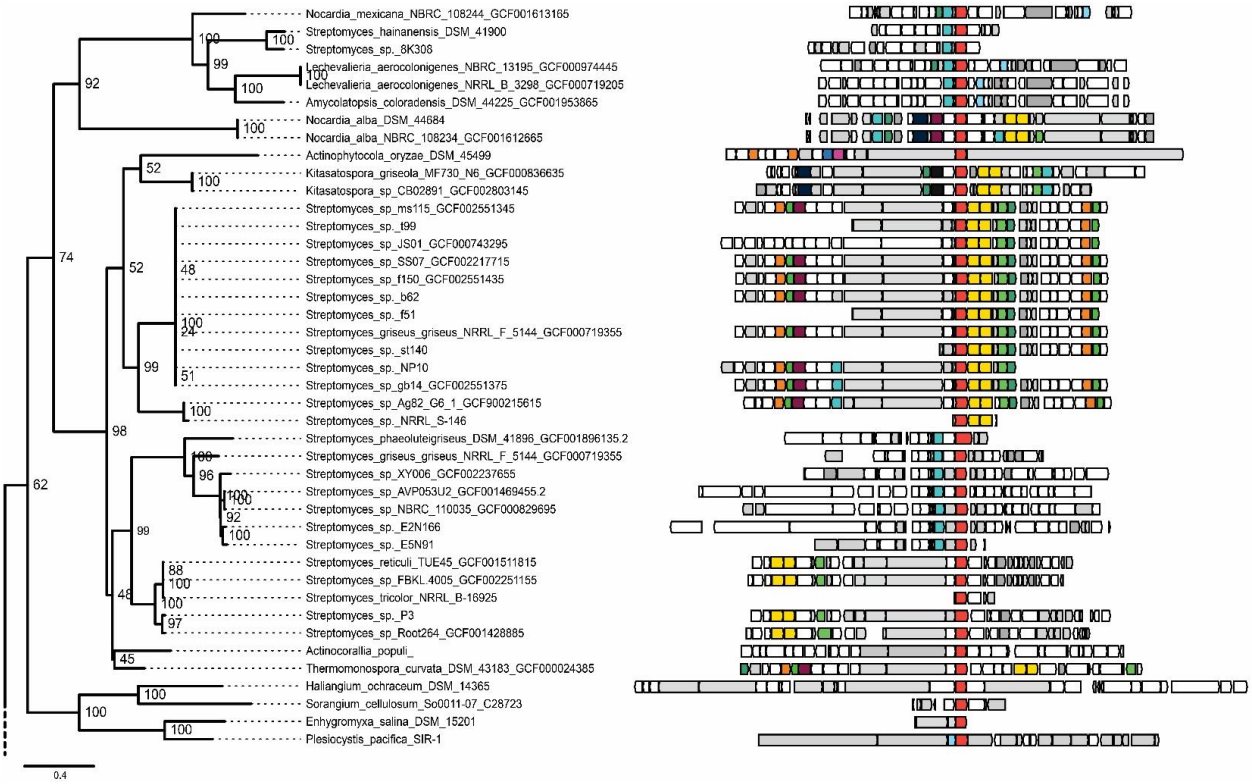


Figure 4-15 Top half of phylogenetic tree and genomic contexts of γ,δ-ACADs identified in this work. This figure represents the top portion of the tree and connects to the bottom portion (Figure 4-16).

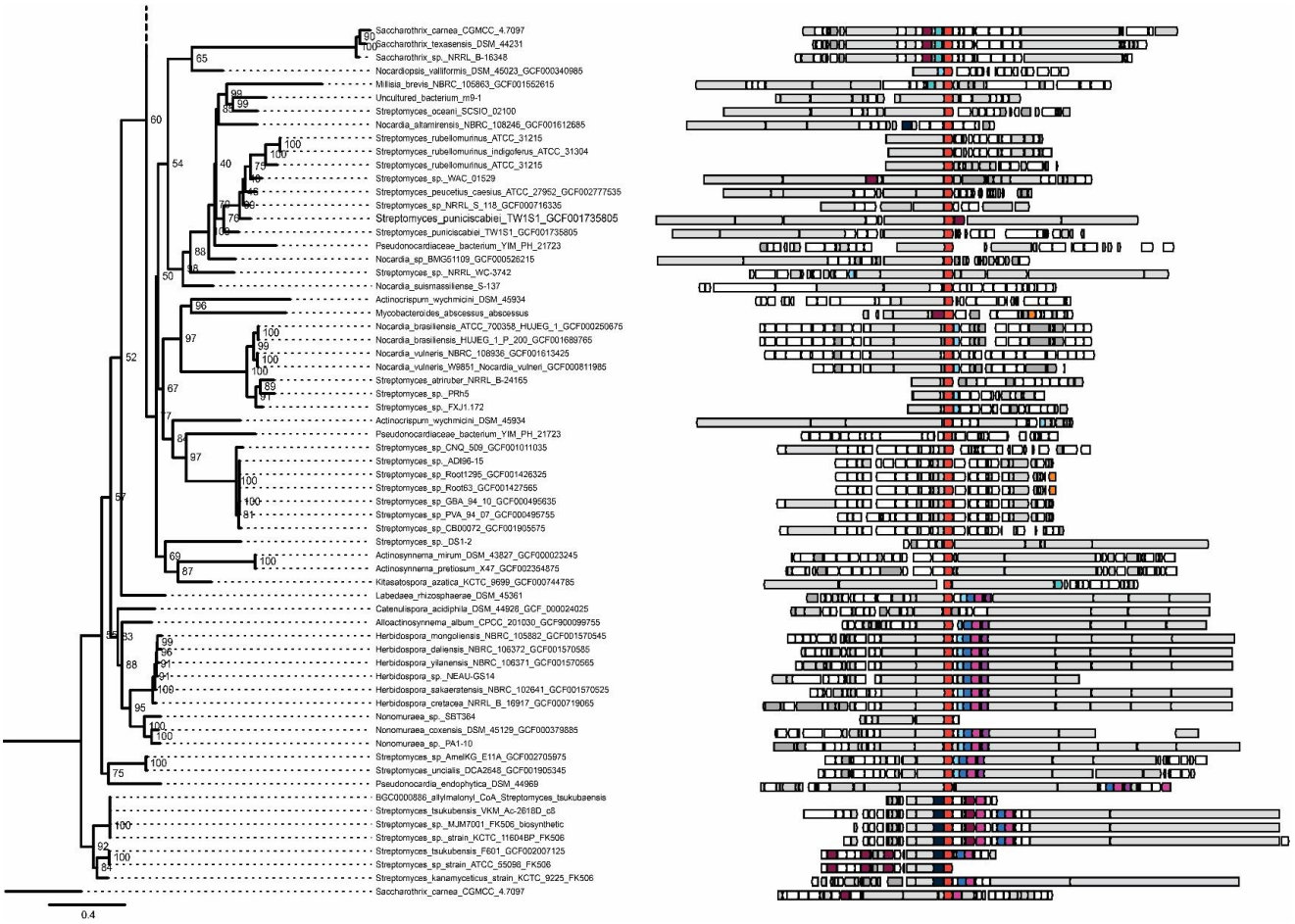


Figure 4-16 Bottom half of phylogenetic tree and genomic contexts of γ,δ-ACADs identified in this work. This figure represents the bottom portion of the tree and connects to the top portion (Figure 4-15).

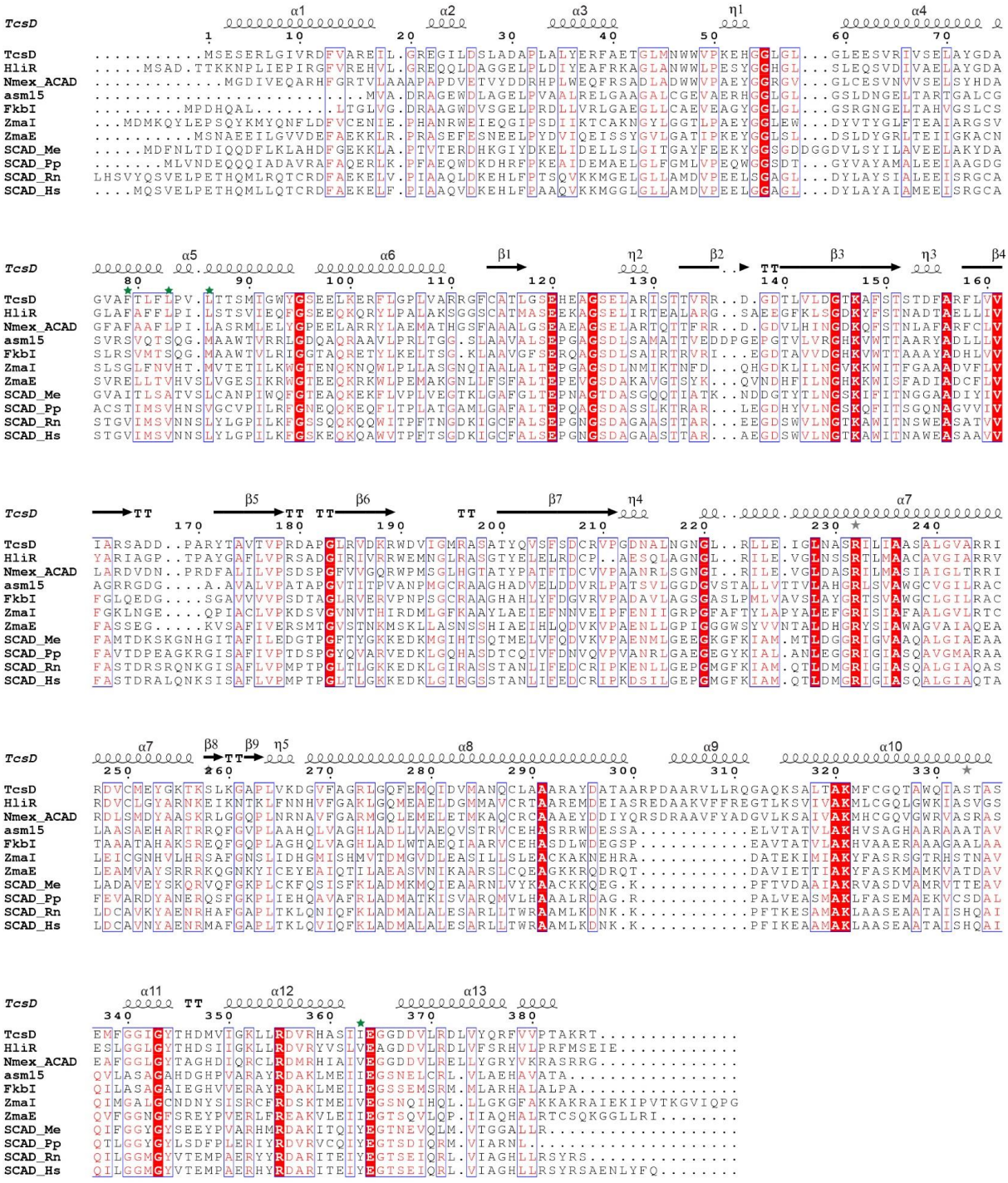
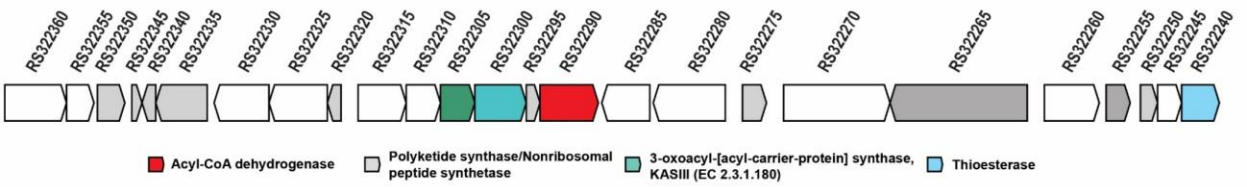


Figure 4-17 Sequence alignment of TcsD with homologs, including Nmex-ACAD. Structural elements are based on the TcsD structure, and green stars denote residues that line the active site pocket. Proteins included in the alignment and their Genbank accession numbers are listed in the description for Figure 4-13.



Locus tag	Gene product	GenBank Protein ID
DFR68_RS32240	thioesterase	WP_068019833.1
DFR68_RS32245	hypothetical protein	WP_068019835.1
DFR68_RS32250	hypothetical protein	WP_068019837.1
DFR68_RS32255	response regulator transcription factor	WP_068019839.1
DFR68_RS32260	histidine kinase	WP_068019841.1
DFR68_RS32265	AAA family ATPase	WP_068019843.1
DFR68_RS32270	MMPL family transporter	WP_068019846.1
DFR68_RS32275	ester cyclase	WP_084519643.1
DFR68_RS32280	MFS transporter	WP_068019850.1
DFR68_RS32285	hypothetical protein	WP_114699758.1
DFR68_RS32290	acyl-CoA dehydrogenase	WP_068019852.1
DFR68_RS32295	acyl carrier protein	WP_068019854.1
DFR68_RS32300	ketoacyl-ACP synthase III	WP_084519644.1
DFR68_RS32305	SDR family oxidoreductase	WP_068019857.1
DFR68_RS32310	HAD family phosphatase	WP_068019859.1
DFR68_RS32315	hypothetical protein	WP_084519645.1
DFR68_RS32320	acyl carrier protein	WP_068019863.1
DFR68_RS32325	beta-ketoacyl-[acyl-carrier-protein] synthase family protein	WP_068019865.1
DFR68_RS32330	hypothetical protein	WP_068019867.1
DFR68_RS32335	DUF2236 domain-containing protein	WP_068019869.1
DFR68_RS32340	hypothetical protein	WP_068019872.1
DFR68_RS32345	hypothetical protein	WP_114699759.1
DFR68_RS32350	SRPBCC family protein	WP_068019876.1
DFR68_RS32355	hypothetical protein	WP_068019878.1
DFR68_RS32360	MCE family protein	WP_084519646.1

Figure 4-18 Genomic context of the *Nocardia mexicana* γ,δ-acyl-CoA dehydrogenase (Nmex-ACAD) and predicted gene products for each coding sequence.

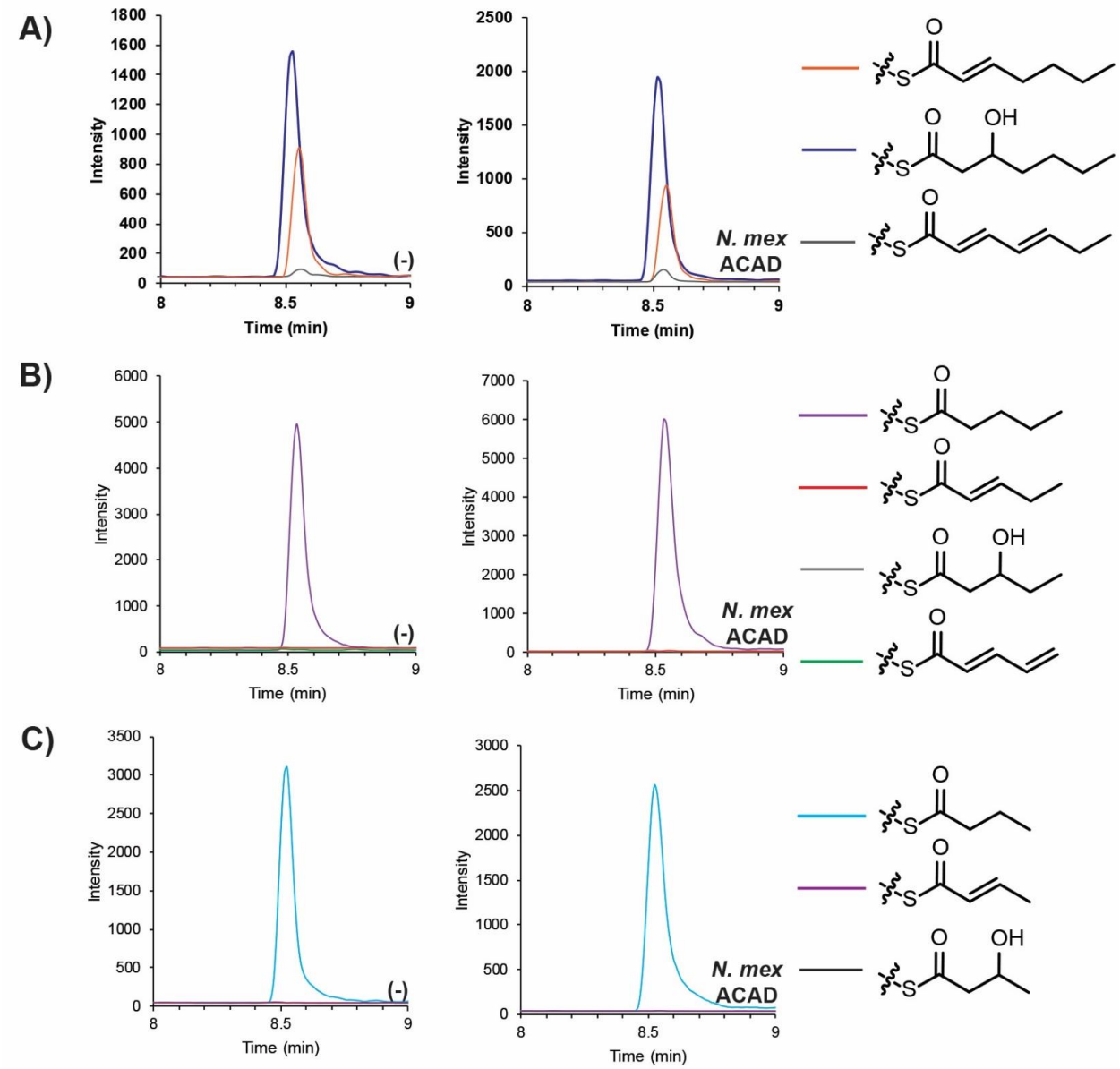


Figure 4-19 Activity of Nmex-ACAD analyzed by targeted LC-MS/MS. *N.mex* ACAD = wild type Nmex-ACAD, (-) = negative control. Chromatograms representing different transitions are color coded (key is shown on right side of figure). The chromatograms shown depict assays on the following substrates: **A)** 2-heptenoyl-ACP, **B)** pentanoyl-ACP, or **C)** butyryl-ACP. For all samples, only the substrate (no product) was observed.

Table 4-1 Uncharacterized homologs of TcsD that contain characteristic γ,δ-ACAD motifs

Species	Protein GenBank Accession	DNA GenBank Accession	Gene cluster type (Antismash annotation)	Closest homologous cluster (Antismash)
Streptomyces kanamyceticus	ADU56239.1	HM116536.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces tacrolimicus	ADU56353.1	HM116538.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces tsukubensis VKM Ac-2618D	WP_006350839.1	NZ_SGFG01000008.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces sp. KCTC 11604BP	ADU56309.1	HM116537.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces tsukubensis NRRL 18488	WP_006350839.1	NZ_AJSZ01000908.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces tsukubensis F601	WP_077974278.1	NZ_MVFC01000056.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Streptomyces sp. MJM7001	WP_006350839.1	HQ696504.1	N/A	Allylmalonyl-ACP/FK506 (T1PKS)
Haliangium ochraceum DSM 14365	AMM72019.1	KU523555.1	N/A	Haliangicin (T1PKS)
Nocardia brasiliensis ATCC 700358	WP_014984668.1	NC_018681.1	T1PKS	Akaeolide (polyketide) (16%)
Nocardia brasiliensis HUJEG-1 isolate P-200	WP_014984668.1	NZ_LRRM01000006.1	T1PKS	Akaeolide (polyketide) (16%)
Saccharomonospora saliphila YIM 90502	WP_019815635.1	NZ_KB912660.1	PKS-like,T1PKS	Arsenopolyketides (45%)

Streptomyces sp. ADI96-15	WP_023416081.1	NZ_ML123109.1	PKS-like,T1PKS	Arsenopolyketides (54%)
Streptomyces sp. PVA 94-07	WP_023416081.1/ESQ07377.1	NZ_CM002273.1	PKS-like,T1PKS	Arsenopolyketides (58%)
Streptomyces sp. Root63	WP_023416081.1	NZ_LMGX01000018.1	PKS-like,T1PKS	Arsenopolyketides (58%)
Streptomyces sp. Root1295	WP_023416081.1	NZ_LMEL01000021.1	PKS-like,T1PKS	Arsenopolyketides (58%)
Streptomyces sp. GBA 94-10	WP_023416081.1	NZ_CM002271.1	PKS-like,T1PKS	Arsenopolyketides (58%)
Streptomyces sp. CB00072	WP_073868671.1	NZ_LIPB01000003.1	PKS-like,T1PKS	Arsenopolyketides (58%)
Millisia brevis	WP_066907456.1	NZ_BCRN01000007.1	NRPS,T1PKS	Aurantimycin (18%)
Pseudonocardia endophytica	WP_132431031.1	NZ_SMFZ01000002.1	NRPS,PKS-like,T1PKS	Butyrolactol A (46%)
Herbidospora sp. NEAU-GS14	WP_137246715.1	NZ_SZQA01000007.1	T1PKS	Butyrolactol A (33%)
Herbidospora yilanensis strain NBRC 106371	WP_062352189.1	NZ_BBXE01000030.1	NRPS,T1PKS	Butyrolactol A (40%)
Streptomyces sp. AmelKG-E11A	WP_099283133.1	NZ_AQRJ01000070.1	T1PKS/NRPS/lassopeptide	Butyrolactol A (40%)
Streptomyces uncialis	WP_073788609.1	NZ_LFBV01000001.1	T1PKS/NRPS/lassopeptide	Butyrolactol A (40%)
Herbidospora cretacea strain NRRL B-16917	WP_030450088.1	NZ_JODQ01000001.1	NRPS,T1PKS,arylpolyene	Butyrolactol A (46%)
Herbidospora daliensis strain NBRC 106372	WP_062432695.1	NZ_BBXF01000003.1	NRPS,T1PKS,arylpolyene	Butyrolactol A (46%)
Herbidospora sakaeratensis strain NBRC 102641	WP_062330499.1	NZ_BBXC01000007.1	NRPS,T1PKS,ladderane	Butyrolactol A (46%)
Herbidospora mongoliensis	WP_066370148.1	NZ_BBXD01000017.1	NRPS,T1PKS,ladderane	Butyrolactol A (46%)

strain NBRC 105882				
Alloactinosynnema album	WP_091377637.1		T1PKS/NRPS	Butyrolactol A (46%)
Nonomuraea sp. PA1-10	WP_139634331	NZ_VDLX01000013.1	NRPS,T1PKS	Butyrolactol A (53%)
Nonomuraea coxensis DSM 45129	WP_020541001.1	NZ_KB903944.1	T1PKS	Butyrolactol A (66%)
Labedaea rhizosphaerae DSM 45361	WP_133849167.1	NZ_SNXZ01000002.1	NRPS,T1PKS,betalactone,transAT-PKS-like	Butyrolactol A (40%)
Nocardia vulneris W9851	WP_052281359.1	NZ_JNFP00000000.1	T1PKS	Chalcomycin (9%)
Nocardia vulneris NBRC 108936	WP_052281359.1	NZ_BDCI01000001.1	T1PKS	Chalcomycin (9%)
Amycolatopsis coloradensis	WP_076160419.1	NZ_MQUQ01000006.1	butyrolactone	Chlorothricin (4%)
Amycolatopsis coloradensis	WP_076160419.1	NZ_MQUQ00000000 (NZ_MQUQ01000006.1)	butyrolactone	Chlorothricin (4%)
Actinocrispum wychmicini	WP_132116054.1	NZ_SLWS01000003.1	PKS-like,T1PKS	Chlorothricin (48%)
Nocardia suismassiliense S-137	WP_107655957.1	NZ_LT985361.1	NRPS,T1PKS,ladderane	Coelimycin (29%)
Nocardia mexicana	WP_068019852.1	NZ_QQAZ01000001	butyrolactone/ladderane	Colabomycin (4%)
Nocardia sp. BMG51109	WP_024802972.1	NZ_JAFQ01000004.1	T1PKS	Stambomycin (40%)
Streptomyces peucetius subsp. caesius ATCC 27952	WP_017584471.1	NZ_CP022438.1	T1PKS,T2PKS	Oligomycin (61%)
Actinosynnema mirum ATCC 29888	WP_015803413.1	NC_013093.1	T1PKS	Cyclizidine (41%)

Streptomyces puniciscabiei	WP_069777509.1	NZ_CP017248.1	T1PKS	E-837 (100%)
Lechevalieria aerocolonigenes	WP_051784425.1	NZ_BBOJ01000031.1	arylpolyene,butyrolactone	Enduracidin (4%)
Saccharothrix sp. NRRL B-16348	WP_053716783.1	NZ_LGED01000125.1	NRPS	Erythrochelin (85%)
Saccharothrix carnea	WP_106615758.1	NZ_PYAX01000004.1	NRPS	Erythrochelin (85%)
Saccharothrix texasensis	WP_123742396.1	NZ_RJKM01000001.1	NRPS	Erythrochelin (85%)
Streptomyces sp. P3	WP_107448820.1	NZ_CP028369.1	LAP,PKS-like,T1PKS,T2PKS	Hedamycin (43%)
Kitasatospora griseola	WP_043910458.1	NZ_JXZB01000002.1	NRPS,PKS-like,T1PKS,T2PKS	Hedamycin (31%)
Kitasatospora sp. CB02891	WP_100586111.1	NZ_NNBO01000004.1	NRPS,PKS-like,T1PKS,T2PKS	Hedamycin (34%)
Streptomyces sp. FBKL.4005	WP_059247715.1	NZ_NPKF01000001.1	LAP,PKS-like,T1PKS,T2PKS	Hedamycin (43%)
Streptomyces reticuli	WP_059247715.1	NZ_LN997842.1	LAP,PKS-like,T2PKS	Hedamycin (46%)
Actinocorallia populi strain A251	WP_106396532.1	NZ_PVZV01000001.1	PKS-like,T1PKS,T2PKS	Hedamycin (56%)
Streptomyces sp. t99	WP_030720926.1	NZ_NTGQ01000069.1	T1PKS,T2PKS	Hedamycin (59%)
Streptomyces sp. st140	WP_030720926.1	NZ_NTGS01000037.1	T2PKS	Hedamycin (59%)
Streptomyces sp. f51	WP_030720926.1	NZ_NTHH01000040.1	T1PKS,T2PKS	Hedamycin (62%)
Thermomonospora curvata ATCC 19995	WP_012853108.1	NC_013510.1	PKS-like,T1PKS,T2PKS	Hedamycin (81%)
Streptomyces sp. ms115	WP_097947834.1	NZ_NTHD01000048.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)
Streptomyces sp. b62	WP_030720926.1	NZ_NTHK01000012.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)

Streptomyces sp. f150	WP_030720926.1	NZ_NTHG01000036.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)
Streptomyces griseus subsp. griseus NRRL F-5144	WP_030720926.1	NZ_JOGA01000026.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)
Streptomyces sp. gb14	WP_030720926.1	NZ_NTHF01000021.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)
Streptomyces sp. SS07	WP_030720926.1	NZ_KZ268499.1	PKS-like,T1PKS,T2PKS,terpene	Hedamycin (87%)
Streptomyces rubellomurinus subsp. indigoferus	KJS54098.1	JZKG00000000.1	T1PKS	Hedamycin (9%)
Streptomyces sp. NRRL S-146	WP_031110854.1	NZ_JOAW01000493.1	T2PKS	Hedamycin (9%)
Streptomyces sp. Ag82_G6-1	WP_097222862.1	NZ_OCNA01000001.1	PKS-like, T1PKS, T2PKS	Hedamycin (90%)
Nocardia alba DSM 44684	TCJ89880.1	SMFR01000008.1	PKS-like,T1PKS,T2PKS	Hedamycin (25%)
Nocardia alba NBRC 108234	WP_067458355.1	NZ_BDAX01000033.1	PKS-like,T1PKS,T2PKS	Hedamycin (25%)
Streptomyces sp. NP10	WP_126932507.1	NZ_PDIQ01000024.1	PKS-like,T1PKS,T2PKS	Hedamycin (50%)
Streptomyces sp. JS01	WP_030720926.1	NZ_JPWW01000020.1	PKS-like,T1PKS,T2PKS	Hedamycin (87%)
Streptomyces sp. Root264	KRD19112.1	NZ_LMIZ01000001	LAP,PKS-like,T1PKS,T2PKS	Hedamycin (43%)
Actinophytocola oryzae DSM 45499	WP_133905254.1	NZ_SOCP01000009.1	T1PKS	Incednine (17%)
Streptomyces sp. WAC 01529	WP_125514820.1	CP029617.1	T1PKS/NRPS	Lorneic acid A (23%)
uncultured bacterium	ASV46999.1	KY560362.1	T1PKS	Lorneic acid A (23%)
Streptomyces oceani	WP_070197325.1	NZ_LJGU00000000.1	T1PKS	Lorneic acid A (23%)
Streptomyces puniciscabiei	WP_069776466.1	NZ_CP017248.1	T1PKS,terpene	Lorneic acid A (28%)

Streptomyces sp. DS1-2	WP_120696069.1	NZ_RBDY01000003.1	T1PKS	Methymycin / pikromycin (57%)
Sorangium cellulosum	KYF78568.1	JEMB01002769.1	T1PKS	Micromonolactam (100%)
Actinosynnema pretiosum strain X47	WP_096495686.1	NZ_CP023445.1	T1PKS	Microtermolide A (33%)
Streptomyces sp. CNQ-509	WP_047018908.1	NZ_CP011492.1	NRPS,PKS-like,T1PKS	Microtermolide A (53%)
Streptomyces sp. AZ1-7	WP_120696069.1	NZ_RBDX01000002.1	T1PKS	Nocardiopsin (21%)
Catenulispora acidiphila DSM 44928	WP_015795553.1	NC_013131.1	T1PKS/NRPS	Octacosamicin (29%)
Streptomyces sp. NRRL S-118	WP_031080613.1	NZ_KL591043.1	T1PKS	Piericidin A1 (58%)
Streptomyces griseus subsp. griseus NRRL F-5144	WP_030723159.1	NZ_JOGA01000040.1	NRPS	Polyoxypeptin (40%)
Streptomyces phaeoluteigriseus strain DSM 41896	OJT46852.1	MPOH00000000.2	NRPS,T1PKS,other	Polyoxypeptin (75%)
Streptomyces sp. XY006	WP_094055126.1	NZ_NOKT01000017.1	NRPS,other	Polyoxypeptin (40%)
Streptomyces sp. E5N91 s-91	WP_121712942.1	NZ_RAIE01000715.1	other	Polyoxypeptin (40%)
Pseudonocardiaceae bacterium YIMPH 21723	WP_120088448.1	NZ_QZFT00000000.1	T1PKS	Pyrronazol B (9%)
Nocardia altamirensis NBRC 108246	WP_069164184.1	NZ_BDAY01000046.1	T1PKS	Stambomycin (40%)
Streptomyces sp. NRRL WC-3742	WP_031071135.1	NZ_JOCF01000031.1	NRPS,T1PKS	Lydicamycin (40%)

Kitasatospora azatica KCTC 9699	WP_083976688.1	NZ_JQMO01000003.1	PKS-like,T1PKS	Zincophorin (61%)
Streptomyces rubellomurinus ATCC 31215	WP_017584471.1	NZ_JZKH01000090.1	N/A	N/A
Streptomyces albidoflavus	WP_128462586.1	NZ_QHCQ00000000.1 (NZ_SCDQ01000064.1)	N/A	N/A
Streptomyces sp. PRh5	WP_051573751.1	NZ_JABQ01000071.1	N/A	N/A
Streptomyces sp. FXJ1.172	WP_067044802.1	NZ_LWRP01000068.1	N/A	N/A
Streptomyces sp. AVP053U2	WP_062189972.1	NZ_LMTQ02000008.1	N/A	N/A
Enhygromyxa salina	KIG11693.1	JMCC00000000.2	N/A	N/A
Streptomyces tricolor NRRL B-16925	WP_086702637.1	NZ_MUMF00000000.1	N/A	N/A
Streptomyces sp. 8K308	WP_132929808.1	NZ_SMKC01000043.1	N/A	N/A
Mycobacteroides abscessus	WP_079869619.1/SHS51726.1	NZ_FSAT01000004.1	N/A	N/A
Streptomyces atriruber	WP_055564739.1	NZ_LIPN01000084.1 (NZ_SMKI00000000.1)	N/A	N/A
Streptomyces hainanensis DSM 41900	WP_132818863.1	NZ_SMKI01000162.1	N/A	N/A
Nocardiopsis valliformis DSM 45023	WP_017584471.1	NZ_ANAZ01000058.1	N/A	N/A
Streptomyces sp. E2N166	WP_121750360.1	NZ_RAIF01000032.1	N/A	N/A
Streptomyces sp. NBRC 110035	WP_042171193.1	NZ_BBNN01000027.1	N/A	N/A

Plesiocystis pacifica SIR-1	WP_006969759.1/EDM81215.1	NZ_ABCS01000005.1	transAT-PKS	N/A
Nonomuraea sp. SBT364	WP_049575489.1	NZ_LAVL01000205.1	T1PKS	N/A

Chapter 5. References

1. Julien, B. *et al.* Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. *Gene* **249,** 153–160 (2000).
2. Bollag, D. M. *et al.* Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55,** 2325–2333 (1995).
3. Luo, G., Pieper, R., Rosa, A., Khosla, C. & Cane, D. E. Erythromycin biosynthesis: exploiting the catalytic versatility of the modular polyketide synthase. *Bioorg. Med. Chem.* **4,** 995–999 (1996).
4. Waldron, C. *et al.* A cluster of genes for the biosynthesis of spinosyns, novel macrolide insect control agents produced by Saccharopolyspora spinosa. *Antonie Van Leeuwenhoek* **78,** 385–390 (2000).
5. Khan, N., Rawlings, B. & Caffrey, P. A labile point in mutant amphotericin polyketide synthases. *Biotechnol. Lett.* **33,** 1121–1126 (2011).
6. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3,** 925–936 (2005).
7. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252,** 675–679 (1991).
8. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat Prod Rep* **29,** 1050–1073 (2012).
9. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat Prod Rep* **27,** 255–278 (2010).
10. Weissman, K. J. The structural basis for docking in modular polyketide biosynthesis. *Chembiochem* **7,** 485–494 (2006).
11. Borgos, S. E. F. *et al.* Probing the structure-function relationship of polyene macrolides: engineered biosynthesis of soluble nystatin analogues. *J. Med. Chem.* **49,** 2431–2439 (2006).
12. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat Prod Rep* **33,** 183–202 (2016).
13. Cummings, M., Breitling, R. & Takano, E. Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiol. Lett.* **351,** 116–125 (2014).

14. Poust, S., Hagen, A., Katz, L. & Keasling, J. D. Narrowing the gap between the promise and reality of polyketide synthases as a synthetic biology platform. *Curr. Opin. Biotechnol.* **30,** 32–39 (2014).
15. Rodriguez, E., Menzella, H. G. & Gramajo, H. in *Complex Enzymes in Microbial Natural Product Biosynthesis, Part B: Polyketides, Aminocoumarins and Carbohydrates* **459,** 339–365 (Elsevier, 2009).
16. Yuzawa, S., Kim, W., Katz, L. & Keasling, J. D. Heterologous production of polyketides by modular type I polyketide synthases in Escherichia coli. *Curr. Opin. Biotechnol.* **23,** 727–735 (2012).
17. Stevens, D. C., Hari, T. P. A. & Boddy, C. N. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat Prod Rep* **30,** 1391–1411 (2013).
18. Long, P. F. *et al.* Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.* **43,** 1215–1225 (2002).
19. Oefner, C., Schulz, H., D'Arcy, A. & Dale, G. E. Mapping the active site of Escherichia coli malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **62,** 613–618 (2006).
20. Dutton, C. J. *et al.* Novel avermectins produced by mutational biosynthesis. *J Antibiot* **44,** 357–365 (1991).
21. Yuzawa, S., Eng, C. H., Katz, L. & Keasling, J. D. Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase. *Biochemistry* **52,** 3791–3793 (2013).
22. Moss, S. J. *et al.* Biosynthesis of the angiogenesis inhibitor borrelidin: directed biosynthesis of novel analogues. *Chem. Commun.* 2341–2343 (2006). doi:10.1039/b602931k
23. Hagen, A. *et al.* In vitro analysis of carboxyacyl substrate tolerance in the loading and first extension modules of borrelidin polyketide synthase. *Biochemistry* **53,** 5975–5977 (2014).
24. August, P. R. *et al.* Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of Amycolatopsis mediterranei S699. *Chem. Biol.* **5,** 69–79 (1998).
25. Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. USA* **92,** 7839–7843 (1995).
26. Traitcheva, N., Jenke-Kodama, H., He, J., Dittmann, E. & Hertweck, C. Non-colinear polyketide biosynthesis in the aureothin and neoaureothin pathways: an evolutionary perspective. *Chembiochem* **8,** 1841–1849 (2007).
27. Trivedi, O. A. *et al.* Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. *Nature* **428,** 441–445 (2004).

28. Motamedi, H. & Shafiee, A. The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur. J. Biochem.* **256,** 528–534 (1998).
29. Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W. & Rao, R. N. Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* **183,** 231–236 (1996).
30. Marsden, A. F. *et al.* Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science* **279,** 199–202 (1998).
31. Rowe, C. J., Cortés, J., Gaisser, S., Staunton, J. & Leadlay, P. F. Construction of new vectors for high-level expression in actinomycetes. *Gene* **216,** 215–223 (1998).
32. Brautaset, T., Borgos, S. E. F., Sletta, H., Ellingsen, T. E. & Zotchev, S. B. Site-specific mutagenesis and domain substitutions in the loading module of the nystatin polyketide synthase, and their effects on nystatin biosynthesis in Streptomyces noursei. *J. Biol. Chem.* **278,** 14913–14919 (2003).
33. Park, S. R., Yoo, Y. J., Ban, Y.-H. & Yoon, Y. J. Biosynthesis of rapamycin and its regulation: past achievements and recent progress. *J Antibiot* **63,** 434–441 (2010).
34. Krastel, P. *et al.* Nannocystin A: an Elongation Factor 1 Inhibitor from Myxobacteria with Differential Anti-Cancer Properties. *Angew. Chem. Int. Ed. Engl.* **54,** 10149–10154 (2015).
35. Heia, S. *et al.* Initiation of polyene macrolide biosynthesis: interplay between polyketide synthase domains and modules as revealed via domain swapping, mutagenesis, and heterologous complementation. *Appl. Environ. Microbiol.* **77,** 6982–6990 (2011).
36. Wilkinson, C. J., Frost, E. J., Staunton, J. & Leadlay, P. F. Chain initiation on the soraphen-producing modular polyketide synthase from Sorangium cellulosum. *Chem. Biol.* **8,** 1197–1208 (2001).
37. Hunziker, D., Yu, T.-W., Hutchinson, C. R., Floss, H. G. & Khosla, C. Primer unit specificity in rifamycin biosynthesis principally resides in the later stages of the biosynthetic pathway. *J. Am. Chem. Soc.* **120,** 1092–1093 (1998).
38. Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E. & Khosla, C. Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. *Biochemistry* **53,** 3796–3806 (2014).
39. Ruan, X. *et al.* Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J. Bacteriol.* **179,** 6416–6425 (1997).
40. Dunn, B. J. & Khosla, C. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *J. R. Soc. Interface* **10,** 20130297 (2013).
41. Lau, J., Fu, H., Cane, D. E. & Khosla, C. Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* **38,** 1643–1651 (1999).

42. Hans, M., Hornung, A., Dziarnowski, A., Cane, D. E. & Khosla, C. Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *J. Am. Chem. Soc.* **125,** 5366–5374 (2003).
43. Koryakina, I. *et al.* Inversion of extender unit selectivity in the erythromycin polyketide synthase by acyltransferase domain engineering. *ACS Chem. Biol.* **12,** 114–123 (2017).
44. Musiol-Kroll, E. M. *et al.* Polyketide bioderivatization using the promiscuous acyltransferase kircii. *ACS Synth. Biol.* **6,** 421–427 (2017).
45. Reeves, C. D. *et al.* Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **40,** 15464–15470 (2001).
46. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* **26,** 90–114 (2009).
47. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14,** 931–943 (2007).
48. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA* **103,** 11124–11129 (2006).
49. Haydock, S. F. *et al.* Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374,** 246–248 (1995).
50. Petković, H. *et al.* Substrate specificity of the acyl transferase domains of EpoC from the epothilone polyketide synthase. *Org. Biomol. Chem.* **6,** 500–506 (2008).
51. Jiang, H. *et al.* An acyltransferase domain of FK506 polyketide synthase recognizing both an acyl carrier protein and coenzyme A as acyl donors to transfer allylmalonyl and ethylmalonyl units. *FEBS J.* **282,** 2527–2539 (2015).
52. Ray, L. & Moore, B. S. Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat Prod Rep* **33,** 150–161 (2016).
53. Sundermann, U. *et al.* Enzyme-directed mutasynthesis: a combined experimental and theoretical approach to substrate recognition of a polyketide synthase. *ACS Chem. Biol.* **8,** 443–450 (2013).
54. Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52,** 1839–1841 (2013).
55. Reeves, C. D. *et al.* Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations[?]. *Biochemistry* **40,** 15464–15470 (2001).

56. Bravo-Rodriguez, K. *et al.* Substrate flexibility of a mutated acyltransferase domain and implications for polyketide biosynthesis. *Chem. Biol.* **22,** 1425–1430 (2015).
57. Koryakina, I., McArthur, J. B., Draelos, M. M. & Williams, G. J. Promiscuity of a modular polyketide synthase towards natural and non-natural extender units. *Org. Biomol. Chem.* **11,** 4449–4458 (2013).
58. Oliynyk, M., Brown, M. J., Cortés, J., Staunton, J. & Leadlay, P. F. A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3,** 833–839 (1996).
59. Liu, L., Thamchaipenet, A., Fu, H., Betlach, M. & Ashley, G. Biosynthesis of 2-Nor-6-deoxyerythronolide B by Rationally Designed Domain Substitution. *J. Am. Chem. Soc.* **119,** 10553–10554 (1997).
60. Stassi, D. L. *et al.* Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc. Natl. Acad. Sci. USA* **95,** 7305–7309 (1998).
61. Petkovic, H. *et al.* A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J Antibiot* **56,** 543–551 (2003).
62. Revill, W. P. *et al.* Genetically engineered analogs of ascomycin for nerve regeneration. *J. Pharmacol. Exp. Ther.* **302,** 1278–1285 (2002).
63. Kato, Y. *et al.* Functional expression of genes involved in the biosynthesis of the novel polyketide chain extension unit, methoxymalonyl-acyl carrier protein, and engineered biosynthesis of 2-desmethyl-2-methoxy-6-deoxyerythronolide B. *J. Am. Chem. Soc.* **124,** 5268–5269 (2002).
64. Yuzawa, S., Kapur, S., Cane, D. E. & Khosla, C. Role of a conserved arginine residue in linkers between the ketosynthase and acyltransferase domains of multimodular polyketide synthases. *Biochemistry* **51,** 3708–3710 (2012).
65. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. USA* **107,** 22066–22071 (2010).
66. Yuzawa, S. *et al.* Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **6,** 139–147 (2017).
67. Walker, M. C. *et al.* Expanding the fluorine chemistry of living systems using engineered polyketide synthase pathways. *Science* **341,** 1089–1094 (2013).
68. Ad, O., Thuronyi, B. W. & Chang, M. C. Y. Elucidating the mechanism of fluorinated extender unit loading for improved production of fluorine-containing polyketides. *Proc. Natl. Acad. Sci. USA* **114,** E660–E668 (2017).

69. Wu, K., Chung, L., Revill, W. P., Katz, L. & Reeves, C. D. The FK520 gene cluster of Streptomyces hygroscopicus var. ascomyceticus (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251,** 81–90 (2000).
70. Ban, Y. H., Park, S. R. & Yoon, Y. J. The biosynthetic pathway of FK506 and its engineering: from past achievements to future prospects. *J. Ind. Microbiol. Biotechnol.* **43,** 389–400 (2016).
71. Chan, Y. A. *et al.* Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units. *Proc. Natl. Acad. Sci. USA* **103,** 14349–14354 (2006).
72. Park, H., Kevany, B. M., Dyer, D. H., Thomas, M. G. & Forest, K. T. A polyketide synthase acyltransferase domain structure suggests a recognition mechanism for its hydroxymalonyl-acyl carrier protein substrate. *PLoS One* **9,** e110965 (2014).
73. Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C. & Cane, D. E. The biochemical basis for stereochemical control in polyketide biosynthesis. *J. Am. Chem. Soc.* **131,** 18501–18511 (2009).
74. Zheng, J., Piasecki, S. K. & Keatinge-Clay, A. T. Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling α-substituent stereochemistry. *ACS Chem. Biol.* **8,** 1964–1971 (2013).
75. Xie, X., Garg, A., Keatinge-Clay, A. T., Khosla, C. & Cane, D. E. Epimerase and reductase activities of polyketide synthase ketoreductase domains utilize the same conserved tyrosine and serine residues. *Biochemistry* **55,** 1179–1186 (2016).
76. Zheng, J., Taylor, C. A., Piasecki, S. K. & Keatinge-Clay, A. T. Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase. *Structure* **18,** 913–922 (2010).
77. Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14,** 898–908 (2007).
78. Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **14,** 737–748 (2006).
79. Zheng, J. & Keatinge-Clay, A. T. Structural and functional analysis of C2-type ketoreductases from modular polyketide synthases. *J. Mol. Biol.* **410,** 105–117 (2011).
80. Bailey, C. B., Pasman, M. E. & Keatinge-Clay, A. T. Substrate structure-activity relationships guide rational engineering of modular polyketide synthase ketoreductases. *Chem. Commun.* **52,** 792–795 (2016).
81. Reid, R. *et al.* A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42,** 72–79 (2003).

82. Bedford, D., Jacobsen, J. R., Luo, G., Cane, D. E. & Khosla, C. A functional chimeric modular polyketide synthase generated via domain replacement. *Chem. Biol.* **3,** 827–831 (1996).
83. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA* **96,** 1846–1851 (1999).
84. Kao, C. M. *et al.* Alcohol Stereochemistry in Polyketide Backbones Is Controlled by the β-Ketoreductase Domains of Modular Polyketide Synthases. *J. Am. Chem. Soc.* **120,** 2478–2479 (1998).
85. McDaniel, R., Kao, C. M., Hwang, S. J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4,** 667–674 (1997).
86. Annaval, T., Paris, C., Leadlay, P. F., Jacob, C. & Weissman, K. J. Evaluating ketoreductase exchanges as a means of rationally altering polyketide stereochemistry. *Chembiochem* **16,** 1357–1364 (2015).
87. Eng, C. H. *et al.* Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type I Modular Polyketide Synthase Subunit. *Biochemistry* **55,** 1677–1680 (2016).
88. Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384,** 941–953 (2008).
89. Kandziora, N. *et al.* Uncovering the origin of Z-configured double bonds in polyketides: intermediate E-double bond formation during borrelidin biosynthesis. *Chem. Sci.* **5,** 3563 (2014).
90. Gay, D., You, Y.-O., Keatinge-Clay, A. & Cane, D. E. Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry* **52,** 8916–8928 (2013).
91. Alhamadsheh, M. M., Palaniappan, N., Daschouduri, S. & Reynolds, K. A. Modular polyketide synthases and cis double bond formation: establishment of activated cis-3-cyclohexylpropenoic acid as the diketide intermediate in phoslactomycin biosynthesis. *J. Am. Chem. Soc.* **129,** 1910–1911 (2007).
92. Akey, D. L. *et al.* Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18,** 94–105 (2010).
93. Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines. *Nat Prod Rep* **33,** 141–149 (2016).
94. Caffrey, P., Lynch, S., Flood, E., Finnan, S. & Oliynyk, M. Amphotericin biosynthesis in Streptomyces nodosus: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* **8,** 713–723 (2001).

95. Ikeda, H., Nonomiya, T., Usami, M., Ohta, T. & Omura, S. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in Streptomyces avermitilis. *Proc. Natl. Acad. Sci. USA* **96,** 9509–9514 (1999).
96. Brautaset, T. *et al.* Biosynthesis of the polyene antifungal antibiotic nystatin in Streptomyces noursei ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. *Chem. Biol.* **7,** 395–403 (2000).
97. Sun, Y. *et al.* A complete gene cluster from Streptomyces nanchangensis NS3226 encoding biosynthesis of the polyether ionophore nanchangmycin. *Chem. Biol.* **10,** 431–441 (2003).
98. Zhou, Y. *et al.* Incomplete beta-ketone processing as a mechanism for polyene structural variation in the FR-008/candicidin complex. *Chem. Biol.* **15,** 629–638 (2008).
99. Brautaset, T. *et al.* Improved antifungal polyene macrolides via engineering of the nystatin biosynthetic genes in Streptomyces noursei. *Chem. Biol.* **15,** 1198–1206 (2008).
100. Yong, J.-H. & Byeon, W.-H. Alternative production of avermectin components in Streptomyces avermitilis by gene replacement. *J. Microbiol.* **43,** 277–284 (2005).
101. Hagen, A. *et al.* Engineering a polyketide synthase for in vitro production of adipic acid. *ACS Synth. Biol.* **5,** 21–27 (2016).
102. Kwan, D. H. *et al.* Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases. *Chem. Biol.* **15,** 1231–1240 (2008).
103. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8,** 615–621 (2012).
104. Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M. & Katz, L. An erythromycin analog produced by reprogramming of polyketide synthesis. *Proc. Natl. Acad. Sci. USA* **90,** 7119–7123 (1993).
105. Faille, A. *et al.* Insights into Substrate Modification by Dehydratases from Type I Polyketide Synthases. *J. Mol. Biol.* **429,** 1554–1569 (2017).
106. Kushnir, S. *et al.* Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library. *Angew. Chem. Int. Ed. Engl.* **51,** 10664–10669 (2012).
107. Kellenberger, L. *et al.* A polylinker approach to reductive loop swaps in modular polyketide synthases. *Chembiochem* **9,** 2740–2749 (2008).
108. Qi, Z. *et al.* Directed accumulation of less toxic pimaricin derivatives by improving the efficiency of a polyketide synthase dehydratase domain. *Appl. Microbiol. Biotechnol.* **101,** 2427–2436 (2017).

109. Zhang, X. *et al.* Construction of ivermectin producer by domain swaps of avermectin polyketide synthase in Streptomyces avermitilis. *Appl. Microbiol. Biotechnol.* **72,** 986–994 (2006).
110. Pinto, A., Wang, M., Horsman, M. & Boddy, C. N. 6-Deoxyerythronolide B synthase thioesterase-catalyzed macrocyclization is highly stereoselective. *Org. Lett.* **14,** 2278–2281 (2012).
111. Kao, C. M., Luo, G., Katz, L., Cane, D. E. & Khosla, C. Engineered biosynthesis of structurally diverse tetraketides by a trimodular polyketide synthase. *J. Am. Chem. Soc.* **118,** 9184–9185 (1996).
112. Kao, C. M. *et al.* Gain of Function Mutagenesis of the Erythromycin Polyketide Synthase. 2. Engineered Biosynthesis of an Eight-Membered Ring Tetraketide Lactone. *J. Am. Chem. Soc.* **119,** 11339–11340 (1997).
113. Sharma, K. K. & Boddy, C. N. The thioesterase domain from the pimaricin and erythromycin biosynthetic pathways can catalyze hydrolysis of simple thioester substrates. *Bioorg. Med. Chem. Lett.* **17,** 3034–3037 (2007).
114. Harper, A. D., Bailey, C. B., Edwards, A. D., Detelich, J. F. & Keatinge-Clay, A. T. Preparative, in vitro biocatalysis of triketide lactone chiral building blocks. *Chembiochem* **13,** 2200–2203 (2012).
115. Akey, D. L. *et al.* Structural basis for macrolactonization by the pikromycin thioesterase. *Nat. Chem. Biol.* **2,** 537–542 (2006).
116. Tripathi, A., Choi, S.-S., Sherman, D. H. & Kim, E.-S. Thioesterase domain swapping of a linear polyketide tautomycetin with a macrocyclic polyketide pikromycin in Streptomyces sp. CK4412. *J. Ind. Microbiol. Biotechnol.* **43,** 1189–1193 (2016).
117. Gehret, J. J. *et al.* Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J. Biol. Chem.* **286,** 14445–14454 (2011).
118. Gu, L. *et al.* Polyketide decarboxylative chain termination preceded by o-sulfonation in curacin a biosynthesis. *J. Am. Chem. Soc.* **131,** 16033–16035 (2009).
119. Dong, Y., Matson, J. B. & Edgar, K. J. Olefin Cross-Metathesis in Polymer and Polysaccharide Chemistry: A Review. *Biomacromolecules* **18,** 1661–1676 (2017).
120. Ciardiello, J. J. *et al.* An expedient strategy for the diversity-oriented synthesis of macrocyclic compounds with natural product-like characteristics. *Tetrahedron* **72,** 3567–3578 (2016).
121. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families : the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol. Life Sci.* **65,** 3895–3906 (2008).

122. Wyatt, M. A., Mok, M. C. Y., Junop, M. & Magarvey, N. A. Heterologous expression and structural characterisation of a pyrazinone natural product assembly line. *Chembiochem* **13,** 2408–2415 (2012).
123. Chhabra, A. *et al.* Nonprocessive [2 + 2]e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. USA* **109,** 5681–5686 (2012).
124. Silakowski, B., Nordsiek, G., Kunze, B., Blöcker, H. & Müller, R. Novel features in a combined polyketide synthase/non-ribosomal peptide synthetase: the myxalamid biosynthetic gene cluster of the myxobacterium Stigmatella aurantiaca Sga1511This article is dedicated to Prof. Dr. E. Leistner on the occasion of his 60th birthday. *Chem. Biol.* **8,** 59–69 (2001).
125. Barajas, J. F. *et al.* Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. *Chem. Biol.* **22,** 1018–1029 (2015).
126. Jin, C., Yao, M., Liu, H., Lee, C. F. & Ji, J. Progress in the production and application of n-butanol as a biofuel. *Renewable and Sustainable Energy Reviews* **15,** 4080–4106 (2011).
127. Dorival, J. *et al.* Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **138,** 4155–4167 (2016).
128. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20,** 1340–1351 (2013).
129. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4,** 41–52 (2009).
130. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284,** 482–485 (1999).
131. Wu, N., Cane, D. E. & Khosla, C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41,** 5056–5066 (2002).
132. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat. Biotechnol.* **23,** 1171–1176 (2005).
133. Klaus, M. *et al.* Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *J. Biol. Chem.* **291,** 16404–16415 (2016).

134. Kennedy, J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat Prod Rep* **25,** 25–34 (2008).
135. Mehlhorn, H., Jones, H. L., Weatherley, A. J. & Schumacher, B. Doramectin, a new avermectin highly efficacious against gastrointestinal nematodes and lungworms of cattle and pigs: two studies carried out under field conditions in Germany. *Parasitol. Res.* **79,** 603–607 (1993).
136. Cropp, T. A., Wilson, D. J. & Reynolds, K. A. Identification of a cyclohexylcarbonyl CoA biosynthetic gene cluster and application in the production of doramectin. *Nat. Biotechnol.* **18,** 980–983 (2000).
137. Zhao, X. *et al.* Construction of a doramectin producer mutant from an avermectin-overproducing industrial strain of Streptomyces avermitilis. *Can. J. Microbiol.* **55,** 1355–1363 (2009).
138. Werneburg, M. *et al.* Exploiting enzymatic promiscuity to engineer a focused library of highly selective antifungal and antiproliferative aureothin analogues. *J. Am. Chem. Soc.* **132,** 10407–10413 (2010).
139. Lowden, P. A. S., Böhm, G. A., Metcalfe, S., Staunton, J. & Leadlay, P. F. New rapamycin derivatives by precursor-directed biosynthesis. *Chembiochem* **5,** 535–538 (2004).
140. Sheehan, L. S. *et al.* Engineering of the spinosyn PKS: directing starter unit incorporation. *J. Nat. Prod.* **69,** 1702–1710 (2006).
141. Lee, H. Y., Harvey, C. J. B., Cane, D. E. & Khosla, C. Improved precursor-directed biosynthesis in E. coli via directed evolution. *J Antibiot* **64,** 59–64 (2011).
142. Kibwage, I. O. *et al.* Identification of novel erythromycin derivatives in mother liquor concentrates of Streptomyces erythraeus. *J Antibiot* **40,** 1–6 (1987).
143. Oliynyk, M. *et al.* Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in Streptomyces cinnamonensis and evidence for the role of monB and monC genes in oxidative cyclization. *Mol. Microbiol.* **49,** 1179–1190 (2003).
144. Bhatt, A. *et al.* Accumulation of an E,E,E-triene by the monensin-producing polyketide synthase when oxidative cyclization is blocked. *Angew. Chem. Int. Ed. Engl.* **44,** 7075–7078 (2005).
145. Bravo-Rodriguez, K. *et al.* Predicted incorporation of non-native substrates by a polyketide synthase yields bioactive natural product derivatives. *Chembiochem* **15,** 1991–1997 (2014).
146. Eustáquio, A. S. *et al.* Biosynthesis of the salinosporamide A polyketide synthase substrate chloroethylmalonyl-coenzyme A from S-adenosyl-L-methionine. *Proc. Natl. Acad. Sci. USA* **106,** 12295–12300 (2009).

147. Eustáquio, A. S. & Moore, B. S. Mutasynthesis of fluorosalinosporamide, a potent and reversible inhibitor of the proteasome. *Angew. Chem. Int. Ed. Engl.* **47,** 3936–3938 (2008).
148. Eustáquio, A. S., O'Hagan, D. & Moore, B. S. Engineering fluorometabolite production: fluorinase expression in Salinispora tropica Yields Fluorosalinosporamide. *J. Nat. Prod.* **73,** 378–382 (2010).
149. Ye, Z., Musiol, E. M., Weber, T. & Williams, G. J. Reprogramming acyl carrier protein interactions of an Acyl-CoA promiscuous trans-acyltransferase. *Chem. Biol.* **21,** 636–646 (2014).
150. Koryakina, I. *et al.* Poly specific trans-acyltransferase machinery revealed via engineered acyl-CoA synthetases. *ACS Chem. Biol.* **8,** 200–208 (2013).
151. Koryakina, I. & Williams, G. J. Mutant malonyl-CoA synthetases with altered specificity for polyketide synthase extender unit generation. *Chembiochem* **12,** 2289–2293 (2011).
152. Rokem, J. S., Lantz, A. E. & Nielsen, J. Systems biology of antibiotic production by microorganisms. *Nat Prod Rep* **24,** 1262–1287 (2007).
153. Olano, C., Lombó, F., Méndez, C. & Salas, J. A. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.* **10,** 281–292 (2008).
154. Li, R. & Townsend, C. A. Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in Streptomyces clavuligerus. *Metab. Eng.* **8,** 240–252 (2006).
155. Zabala, D., Braña, A. F., Flórez, A. B., Salas, J. A. & Méndez, C. Engineering precursor metabolite pools for increasing production of antitumor mithramycins in Streptomyces argillaceus. *Metab. Eng.* **20,** 187–197 (2013).
156. Zhang, H., Boghigian, B. A. & Pfeifer, B. A. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnol. Bioeng.* **105,** 567–573 (2010).
157. Vandova, G. A. *et al.* Heterologous expression of diverse propionyl-CoA carboxylases affects polyketide production in Escherichia coli. *J Antibiot* **70,** 859–863 (2017).
158. Arabolaza, A. *et al.* Crystal structures and mutational analyses of acyl-CoA carboxylase beta subunit of Streptomyces coelicolor. *Biochemistry* **49,** 7367–7376 (2010).
159. Ray, L. *et al.* A crotonyl-CoA reductase-carboxylase independent pathway for assembly of unusual alkylmalonyl-CoA polyketide synthase extender units. *Nat. Commun.* **7,** 13609 (2016).

160. Rodriguez, E. *et al.* Engineered biosynthesis of 16-membered macrolides that require methoxymalonyl-ACP precursors in Streptomyces fradiae. *Appl. Microbiol. Biotechnol.* **66,** 85–91 (2004).
161. Lu, C., Zhang, X., Jiang, M. & Bai, L. Enhanced salinomycin production by adjusting the supply of polyketide extender units in Streptomyces albus. *Metab. Eng.* **35,** 129–137 (2016).
162. Bunet, R. *et al.* A single Sfp-type phosphopantetheinyl transferase plays a major role in the biosynthesis of PKS and NRPS derived metabolites in Streptomyces ambofaciens ATCC23877. *PLoS One* **9,** e87607 (2014).
163. Chen, X.-H. *et al.* Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. *J. Bacteriol.* **188,** 4024–4036 (2006).
164. Volokhan, O., Sletta, H., Sekurova, O. N., Ellingsen, T. E. & Zotchev, S. B. An unexpected role for the putative 4'-phosphopantetheinyl transferase-encoding gene nysF in the regulation of nystatin biosynthesis in Streptomyces noursei ATCC 11455. *FEMS Microbiol. Lett.* **249,** 57–64 (2005).
165. Beld, J., Sonnenschein, E. C., Vickery, C. R., Noel, J. P. & Burkart, M. D. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat Prod Rep* **31,** 61–108 (2014).
166. Quadri, L. E. *et al.* Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37,** 1585–1595 (1998).
167. Peypoux, F., Bonmatin, J. M. & Wallach, J. Recent trends in the biochemistry of surfactin. *Appl. Microbiol. Biotechnol.* **51,** 553–563 (1999).
168. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* **291,** 1790–1792 (2001).
169. Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R. & Kealey, J. T. Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production. *J. Ind. Microbiol. Biotechnol.* **30,** 500–509 (2003).
170. Mutka, S. C., Bondi, S. M., Carney, J. R., Da Silva, N. A. & Kealey, J. T. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. *FEMS Yeast Res* **6,** 40–47 (2006).
171. Kealey, J. T., Liu, L., Santi, D. V., Betlach, M. C. & Barr, P. J. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc. Natl. Acad. Sci. USA* **95,** 505–509 (1998).
172. Mootz, H. D., SchÃ¶rgendorfer, K. & Marahiel, M. A. Functional characterization of 4â®²-phosphopantetheinyl transferase genes of bacterial and fungal origin by

complementation of*Saccharomyces cerevisiae lys5*. *FEMS Microbiol. Lett.* **213,** 51–57 (2002).
173. Zobel, S., Kumpfmüller, J., Süssmuth, R. D. & Schweder, T. Bacillus subtilis as heterologous host for the secretory production of the non-ribosomal cyclodepsipeptide enniatin. *Appl. Microbiol. Biotechnol.* **99,** 681–691 (2015).
174. Kumpfmüller, J. *et al.* Production of the polyketide 6-deoxyerythronolide B in the heterologous host Bacillus subtilis. *Appl. Microbiol. Biotechnol.* **100,** 1209–1220 (2016).
175. Sánchez, C., Du, L., Edwards, D. J., Toney, M. D. & Shen, B. Cloning and characterization of a phosphopantetheinyl transferase from Streptomyces verticillus ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin. *Chem. Biol.* **8,** 725–738 (2001).
176. Baltz, R. H. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.* **37,** 759–772 (2010).
177. Jiang, H. *et al.* Improvement of natamycin production by engineering of phosphopantetheinyl transferases in Streptomyces chattanoogensis L10. *Appl. Environ. Microbiol.* **79,** 3346–3354 (2013).
178. Zhang, B. *et al.* Activation of natural products biosynthetic pathways via a protein modification level regulation. *ACS Chem. Biol.* **12,** 1732–1736 (2017).
179. Stratigopoulos, G., Gandecha, A. R. & Cundliffe, E. Regulation of tylosin production and morphological differentiation in Streptomyces fradiae by TylP, a deduced gamma-butyrolactone receptor. *Mol. Microbiol.* **45,** 735–744 (2002).
180. Stratigopoulos, G., Bate, N. & Cundliffe, E. Positive control of tylosin biosynthesis: pivotal role of TylR. *Mol. Microbiol.* **54,** 1326–1334 (2004).
181. Garcia-Bernardo, J., Braña, A. F., Méndez, C. & Salas, J. A. Insertional inactivation of mtrX and mtrY genes from the mithramycin gene cluster affects production and growth of the producer organism Streptomyces argillaceus. *FEMS Microbiol. Lett.* **186,** 61–65 (2000).
182. Lombó, F., Braña, A. F., Méndez, C. & Salas, J. A. The mithramycin gene cluster of Streptomyces argillaceus contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster. *J. Bacteriol.* **181,** 642–647 (1999).
183. Song, K. *et al.* Engineering of the LysR family transcriptional regulator FkbR1 and its target gene to improve ascomycin production. *Appl. Microbiol. Biotechnol.* **101,** 4581–4592 (2017).
184. Stratigopoulos, G. & Cundliffe, E. Expression analysis of the tylosin-biosynthetic gene cluster: pivotal regulatory role of the tylQ product. *Chem. Biol.* **9,** 71–78 (2002).

185. Shao, Z. *et al.* Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol.* **2,** 662–669 (2013).
186. Basitta, P. *et al.* AGOS: A Plug-and-Play Method for the Assembly of Artificial Gene Operons into Functional Biosynthetic Gene Clusters. *ACS Synth. Biol.* **6,** 817–825 (2017).
187. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096 (2014).
188. Zhang, M. M., Wang, Y., Ang, E. L. & Zhao, H. Engineering microbial hosts for production of bacterial natural products. *Nat Prod Rep* **33,** 963–987 (2016).
189. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin. (Shanghai)* **47,** 231–243 (2015).
190. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4,** 723–728 (2015).
191. Zhang, M. M. *et al.* CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nat. Chem. Biol.* (2017). doi:10.1038/nchembio.2341
192. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A. & Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu. Rev. Biochem.* **76,** 195–221 (2007).
193. Yan, J., Hazzard, C., Bonnett, S. A. & Reynolds, K. A. Functional modular dissection of DEBS1-TE changes triketide lactone ratios and provides insight into Acyl group loading, hydrolysis, and ACP transfer. *Biochemistry* **51,** 9333–9341 (2012).
194. Kao, C. M., Katz, L. & Khosla, C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265,** 509–512 (1994).
195. Hertweck, C. Decoding and reprogramming complex polyketide assembly lines: prospects for synthetic biology. *Trends Biochem. Sci.* **40,** 189–199 (2015).
196. Licona-Cassani, C. *et al.* Systems biology approaches to understand natural products biosynthesis. *Front. Bioeng. Biotechnol.* **3,** 199 (2015).
197. Pérez-Victoria, I., Martín, J. & Reyes, F. Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta Med.* **82,** 857–871 (2016).
198. Covington, B. C., McLean, J. A. & Bachmann, B. O. Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Nat Prod Rep* **34,** 6–24 (2017).
199. Hoffmann, T., Krug, D., Hüttel, S. & Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **86,** 10780–10788 (2014).

200. Rogers, J. K. *et al.* Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* **43,** 7648–7660 (2015).
201. Tremblay, N., Hill, P., Conway, K. R. & Boddy, C. N. The use of clustermine360 for the analysis of polyketide and nonribosomal peptide biosynthetic pathways. *Methods Mol. Biol.* **1401,** 233–252 (2016).
202. Khater, S. *et al.* SBSPKSv2: structure-based sequence analysis of polyketide synthases and non-ribosomal peptide synthetases. *Nucleic Acids Res.* **45,** W72–W79 (2017).
203. Ichikawa, N. *et al.* DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **41,** D408–14 (2013).
204. Blin, K. *et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45,** W36–W41 (2017).
205. Conway, K. R. & Boddy, C. N. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. *Nucleic Acids Res.* **41,** D402–7 (2013).
206. Medema, M. H. *et al.* Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **11,** 625–631 (2015).
207. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661 (2016).
208. Dias, D. A., Urban, S. & Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2,** 303–336 (2012).
209. Scott, T. A. & Piel, J. The hidden enzymology of bacterial natural product biosynthesis. *Nat. Rev. Chem.* (2019). doi:10.1038/s41570-019-0107-1
210. Minowa, Y., Araki, M. & Kanehisa, M. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes. *J. Mol. Biol.* **368,** 1500–1517 (2007).
211. Waldron, C. *et al.* Cloning and analysis of the spinosad biosynthetic gene cluster of Saccharopolyspora spinosa. *Chem. Biol.* **8,** 487–499 (2001).
212. Copéret, C., Adolfsson, H. & Barry Sharpless, K. A simple and efficient method for epoxidation of terminal alkenes. *Chem. Commun.* 1565–1566 (1997). doi:10.1039/a703542j
213. Mlynarski, S. N., Schuster, C. H. & Morken, J. P. Asymmetric synthesis from terminal alkenes by cascades of diboration and cross-coupling. *Nature* **505,** 386–390 (2014).
214. Becker, H. & Sharpless, K. B. A new ligand class for the asymmetric dihydroxylation of olefins. *Angew. Chem. Int. Ed. Engl.* **35,** 448–451 (1996).
215. Xu, J. *et al.* Copper-catalyzed trifluoromethylation of terminal alkenes through allylic C-H bond activation. *J. Am. Chem. Soc.* **133,** 15300–15303 (2011).

216. Breit, B. & Seiche, W. Hydrogen bonding as a construction element for bidentate donor ligands in homogeneous catalysis: regioselective hydroformylation of terminal alkenes. *J. Am. Chem. Soc.* **125,** 6608–6609 (2003).
217. Urkalan, K. B. & Sigman, M. S. Palladium-catalyzed oxidative intermolecular difunctionalization of terminal alkenes with organostannanes and molecular oxygen. *Angew. Chem. Int. Ed. Engl.* **48,** 3146–3149 (2009).
218. Rucker, R. P., Whittaker, A. M., Dang, H. & Lalic, G. Synthesis of tertiary alkyl amines from terminal alkenes: copper-catalyzed amination of alkyl boranes. *J. Am. Chem. Soc.* **134,** 6571–6574 (2012).
219. Coombs, J. R. & Morken, J. P. Catalytic enantioselective functionalization of unactivated terminal alkenes. *Angew. Chem. Int. Ed. Engl.* **55,** 2636–2649 (2016).
220. Nishimura, T., Kakiuchi, N., Onoue, T., Ohe, K. & Uemura, S. Palladium(II)-catalyzed oxidation of terminal alkenes to methyl ketones using molecular oxygen. *J. Chem. Soc., Perkin Trans. 1* 1915–1918 (2000). doi:10.1039/b001854f
221. Jensen, C. M. & Trogler, W. C. Catalytic hydration of terminal alkenes to primary alcohols. *Science* **233,** 1069–1071 (1986).
222. Du, X. & Huang, Z. Advances in Base-Metal-Catalyzed Alkene Hydrosilylation. *ACS Catal.* **7,** 1227–1243 (2017).
223. Fürstner, A. Olefin metathesis and beyond. *Angew. Chem. Int. Ed.* **39,** 3012–3043 (2000).
224. Buslov, I., Becouse, J., Mazza, S., Montandon-Clerc, M. & Hu, X. Chemoselective alkene hydrosilylation catalyzed by nickel pincer complexes. *Angew. Chem. Int. Ed. Engl.* **54,** 14523–14526 (2015).
225. Cannon, J. S. & Grubbs, R. H. Alkene chemoselectivity in ruthenium-catalyzed Z-selective olefin metathesis. *Angew. Chem. Int. Ed. Engl.* **52,** 9001–9004 (2013).
226. Cristina Silva Costa, D. Additions to non-activated alkenes: Recent advances. *Arabian Journal of Chemistry* (2017). doi:10.1016/j.arabjc.2017.07.017
227. Chatterjee, A. K., Choi, T.-L., Sanders, D. P. & Grubbs, R. H. A general model for selectivity in olefin cross metathesis. *J. Am. Chem. Soc.* **125,** 11360–11370 (2003).
228. O'Leary, D. J., Blackwell, H. E., Washenfelder, R. A. & Grubbs, R. H. A new method for cross-metathesis of terminal olefins. *Tetrahedron Lett.* **39,** 7427–7430 (1998).
229. Andrus, M. B., Lepore, S. D. & Sclafani, J. A. Selective dihydroxylation of non-conjugated dienes in favor of the terminal olefin. *Tetrahedron Lett.* **38,** 4043–4046 (1997).
230. Campbell, A. N., White, P. B., Guzei, I. A. & Stahl, S. S. Allylic C-H acetoxylation with a 4,5-diazafluorenone-ligated palladium catalyst: a ligand-based strategy to achieve aerobic catalytic turnover. *J. Am. Chem. Soc.* **132,** 15116–15119 (2010).

231. Bigi, M. A. & White, M. C. Terminal olefins to linear α,β-unsaturated ketones: Pd(II)/hypervalent iodine co-catalyzed Wacker oxidation-dehydrogenation. *J. Am. Chem. Soc.* **135,** 7831–7834 (2013).
232. Dong, G., Teo, P., Wickens, Z. K. & Grubbs, R. H. Primary alcohols from terminal olefins: formal anti-Markovnikov hydration via triple relay catalysis. *Science* **333,** 1609–1612 (2011).
233. Klar, U. *et al.* Total synthesis and antitumor activity of ZK-EPO: the first fully synthetic epothilone in clinical development. *Angew. Chem. Int. Ed. Engl.* **45,** 7942–7948 (2006).
234. Klar, U., Hoffmann, J. & Giurescu, M. Sagopilone (ZK-EPO): from a natural product to a fully synthetic clinical development candidate. *Expert Opin. Investig. Drugs* **17,** 1735–1748 (2008).
235. Kino, T. *et al.* FK-506, a novel immunosuppressant isolated from a Streptomyces. I. Fermentation, isolation, and physico-chemical and biological characteristics. *J Antibiot* **40,** 1249–1255 (1987).
236. Mitsuhashi, S. *et al.* Tautomycetin is a novel and specific inhibitor of serine/threonine protein phosphatase type 1, PP1. *Biochem. Biophys. Res. Commun.* **287,** 328–331 (2001).
237. Fudou, R., Iizuka, T. & Yamanaka, S. Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 1. Fermentation and biological characteristics. *J Antibiot* **54,** 149–152 (2001).
238. Scaglione, J. B. *et al.* Biochemical and structural characterization of the tautomycetin thioesterase: analysis of a stereoselective polyketide hydrolase. *Angew. Chem. Int. Ed. Engl.* **49,** 5726–5730 (2010).
239. Li, W. *et al.* Characterization of the tautomycetin biosynthetic gene cluster from Streptomyces griseochromogenes provides new insight into dialkylmaleic anhydride biosynthesis. *J. Nat. Prod.* **72,** 450–459 (2009).
240. Sun, Y. *et al.* Heterologous production of the marine myxobacterial antibiotic haliangicin and its unnatural analogues generated by engineering of the biochemical pathway. *Sci. Rep.* **6,** 22091 (2016).
241. Mo, S. *et al.* Biosynthesis of the allylmalonyl-CoA extender unit for the FK506 polyketide synthase proceeds through a dedicated polyketide synthase and facilitates the mutasynthesis of analogues. *J. Am. Chem. Soc.* **133,** 976–985 (2011).
242. Ghisla, S. & Thorpe, C. Acyl-CoA dehydrogenases. A mechanistic overview. *Eur. J. Biochem.* **271,** 494–508 (2004).
243. Fendrich, G. & Abeles, R. H. Mechanism of action of butyryl-CoA dehydrogenase: reactions with acetylenic, olefinic, and fluorinated substrate analogues. *Biochemistry* **21,** 6685–6695 (1982).

244. Powell, P. J. & Thorpe, C. 2-octynoyl coenzyme A is a mechanism-based inhibitor of pig kidney medium-chain acyl coenzyme A dehydrogenase: isolation of the target peptide. *Biochemistry* **27,** 8022–8028 (1988).
245. Kim, J. J., Wang, M. & Paschke, R. Crystal structures of medium-chain acyl-CoA dehydrogenase from pig liver mitochondria with and without substrate. *Proc. Natl. Acad. Sci. USA* **90,** 7523–7527 (1993).
246. Ghisla, S., Thorpe, C. & Massey, V. Mechanistic studies with general acyl-CoA dehydrogenase and butyryl-CoA dehydrogenase: evidence for the transfer of the beta-hydrogen to the flavin N(5)-position as a hydride. *Biochemistry* **23,** 3154–3161 (1984).
247. Trievel, R. C., Wang, R., Anderson, V. E. & Thorpe, C. Role of the carbonyl group in thioester chain length recognition by the medium chain acyl-CoA dehydrogenase. *Biochemistry* **34,** 8597–8605 (1995).
248. Mancini-Samuelson, G. J., Kieweg, V., Sabaj, K. M., Ghisla, S. & Stankovich, M. T. Redox properties of human medium-chain acyl-CoA dehydrogenase, modulation by charged active-site amino acid residues. *Biochemistry* **37,** 14605–14612 (1998).
249. Becker, D. F., Fuchs, J. A. & Stankovich, M. T. Product binding modulates the thermodynamic properties of a Megasphaera elsdenii short-chain acyl-CoA dehydrogenase active-site mutant. *Biochemistry* **33,** 7082–7087 (1994).
250. Vock, P., Engst, S., Eder, M. & Ghisla, S. Substrate activation by acyl-CoA dehydrogenases: transition-state stabilization and pKs of involved functional groups. *Biochemistry* **37,** 1848–1860 (1998).
251. Engst, S., Vock, P., Wang, M., Kim, J. J. & Ghisla, S. Mechanism of activation of acyl-CoA substrates by medium chain acyl-CoA dehydrogenase: interaction of the thioester carbonyl with the flavin adenine dinucleotide ribityl side chain. *Biochemistry* **38,** 257–267 (1999).
252. Lau, S. M., Brantley, R. K. & Thorpe, C. The reductive half-reaction in Acyl-CoA dehydrogenase from pig kidney: studies with thiaoctanoyl-CoA and oxaoctanoyl-CoA analogues. *Biochemistry* **27,** 5089–5095 (1988).
253. Stankovich, M. T., Sabaj, K. M. & Tonge, P. J. Structure/function of medium chain acyl-CoA dehydrogenase: the importance of substrate polarization. *Arch. Biochem. Biophys.* **370,** 16–21 (1999).
254. Lehman, T. C., Hale, D. E., Bhala, A. & Thorpe, C. An acyl-coenzyme a dehydrogenase assay utilizing the ferricenium ion. *Anal. Biochem.* **186,** 280–284 (1990).
255. Dorrestein, P. C. *et al.* Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* **45,** 12756–12766 (2006).

256. Cummings, J. G. & Thorpe, C. 3-Methyleneoctanoyl-CoA and 3-methyl-trans-2-octenoyl-CoA: two new mechanism-based inhibitors of medium chain acyl-CoA dehydrogenase from pig kidney. *Biochemistry* **33,** 788–797 (1994).
257. Satoh, A. *et al.* Structure of the transition state analog of medium-chain acyl-CoA dehydrogenase. Crystallographic and molecular orbital studies on the charge-transfer complex of medium-chain acyl-CoA dehydrogenase with 3-thiaoctanoyl-CoA. *J. Biochem.* **134,** 297–304 (2003).
258. Battaile, K. P. *et al.* Crystal structure of rat short chain acyl-CoA dehydrogenase complexed with acetoacetyl-CoA: comparison with other acyl-CoA dehydrogenases. *J. Biol. Chem.* **277,** 12200–12207 (2002).
259. Djordjevic, S., Pace, C. P., Stankovich, M. T. & Kim, J. J. Three-dimensional structure of butyryl-CoA dehydrogenase from Megasphaera elsdenii. *Biochemistry* **34,** 2163–2171 (1995).
260. Watanabe, K., Khosla, C., Stroud, R. M. & Tsai, S. C. Crystal structure of an Acyl-ACP dehydrogenase from the FK520 polyketide biosynthetic pathway: insights into extender unit biosynthesis. *J. Mol. Biol.* **334,** 435–444 (2003).
261. McMahon, B., Gallagher, M. E. & Mayhew, S. G. The protein coded by the PP2216 gene of Pseudomonas putida KT2440 is an acyl-CoA dehydrogenase that oxidises only short-chain aliphatic substrates. *FEMS Microbiol. Lett.* **250,** 121–127 (2005).
262. Cruz-Morales, P. *et al.* Actinobacteria phylogenomics, selective isolation from an iron oligotrophic environment and siderophore functional characterization, unveil new desferrioxamine traits. *FEMS Microbiol. Ecol.* **93,** (2017).
263. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
264. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39,** W29–37 (2011).
265. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345 (2009).
266. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4,** e5553 (2009).
267. Oberortner, E., Cheng, J.-F., Hillson, N. J. & Deutsch, S. Streamlining the Design-to-Build Transition with Build-Optimization Software Tools. *ACS Synth. Biol.* **6,** 485–496 (2017).
268. Jancarik, J. & Kim, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* **24,** 409–411 (1991).
269. McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).

270. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **68,** 352–367 (2012).
271. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **60,** 2126–2132 (2004).
272. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66,** 213–221 (2010).
273. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–83 (2007).
274. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.' ' . at <https://pymol.org/2/>
275. Mishra, P. K. & Drueckhammer, D. G. Coenzyme A Analogues and Derivatives: Synthesis and Applications as Mechanistic Probes of Coenzyme A Ester-Utilizing Enzymes. *Chem. Rev.* **100,** 3283–3310 (2000).
276. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–143 (1984).
277. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26,** 966–968 (2010).
278. Sharma, V. *et al.* Panorama public: A public repository for quantitative data sets processed in skyline. *Mol. Cell Proteomics* **17,** 1239–1244 (2018).
279. Baidoo, E. E. K., Wang, G., Joshua, C. J., Benites, V. T. & Keasling, J. D. Liquid Chromatography and Mass Spectrometry Analysis of Isoprenoid Intermediates in Escherichia coli. *Methods Mol. Biol.* **1859,** 209–224 (2019).
280. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39,** W339–46 (2011).
281. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539 (2011).
282. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42,** W320–4 (2014).
283. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33,** 511–518 (2005).
284. Okonechnikov, K., Golosova, O., Fursov, M. & UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28,** 1166–1167 (2012).
285. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics* **27,** 1575–1577 (2011).

286. Bakan, A. *et al.* Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **30,** 2681–2683 (2014).
287. Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W. & Cohn, J. D. Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **62,** 915–922 (2006).
288. Terwilliger, T. C., Adams, P. D., Moriarty, N. W. & Cohn, J. D. Ligand identification using electron-density map correlations. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **63,** 101–107 (2007).
289. Lee, T.S., Krupa, R.A., Zhang, F., Hajimorad, M., Holtz, W.J., Prasad, N., Lee, S.K., Keasling, J.D., BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng* **5**, 12 (2011).